```
                    GenCore version 4.5
              Copyright (c) 1993 - 2000 Compugen Ltd.


OM nucleic - nucleic search, using sw model

Run on:        November 7, 2001, 03:28:25 ; Search time 1190.53 Seconds
                                            (without alignments)
                                            6041.430 Million cell updates/sec

Title:         US-09-521-640-2
Perfect score: 465
Sequence:      1 ggggagtatcgtgagttcac.........cccttcttgttcaaaacacn 465

Scoring table: OLIGO_NUC
               Gapop 60.0 , Gapext 60.0

Se___ned:      1344157 seqs, 7733874588 residues

Word size :    30

Total number of hits satisfying chosen parameters:       0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :      GenEmbl:*
          1:  gb_ba1:*          44:  em_ov:*
          2:  gb_ba2:*          45:  em_pat:*
          3:  gb_ba3:*          46:  em_ph:*
          4:  gb_in1:*          47:  em_pl:*
          5:  gb_in2:*          48:  em_ro:*
          6:  gb_in3:*          49:  em_sts:*
          7:  gb_om:*           50:  em_sy:*
          8:  gb_ov:*           51:  em_un:*
          9:  gb_pat1:*         52:  em_vi:*
         10:  gb_pat2:*         53:  gb_sts1:*
         11:  gb_ph:*           54:  gb_sts2:*
         12:  gb_pl1:*          55:  gb_sts3:*
         13:  gb_pl2:*          56:  gb_sy:*
         14:  gb_pl3:*          57:  gb_un:*
         15:  gb_pl4:*          58:  gb_vil:*
         16:  em_ba1:*          59:  gb_vi2:*
         17:  em_ba2:*          60:  gb_htg1:*
         18:  em_fun:*          61:  gb_htg2:*
         19:  em_htgo_hum:*     62:  gb_htg3:*
         20:  em_htgo_inv:*     63:  gb_htg4:*
         21:  em_htgo_rod:*     64:  gb_htg5:*
         22:  em_htg_hum1:*     65:  gb_htg6:*
         23:  em_htg_hum2:*     66:  gb_htg7:*
         24:  em_htg_hum3:*     67:  gb_htg8:*
         25:  em_htg_hum4:*     68:  gb_htg9:*
         26:  em_htg_hum5:*     69:  gb_htg10:*
         27:  em_htg_hum6:*     70:  gb_htg11:*
         28:  em_htg_hum7:*     71:  gb_htg12:*
         29:  em_htg_hum8:*     72:  gb_htg13:*
         30:  em_htg_inv1:*     73:  gb_htg14:*
         31:  em_htg_inv2:*     74:  gb_htg15:*
         32:  em_htg_other:*    75:  gb_htg16:*
         33:  em_htg_rod:*      76:  gb_htg17:*
         34:  em_hum1:*         77:  gb_htg18:*
         35:  em_hum2:*         78:  gb_htg19:*
         36:  em_hum3:*         79:  gb_htg20:*
         37:  em_hum4:*         80:  gb_htg21:*
         38:  em_hum5:*         81:  gb_htg22:*
         39:  em_hum6:*         82:  gb_htg23:*
         40:  em_hum7:*         83:  gb_htg24:*
         41:  em_in:*           84:  gb_htg25:*
         42:  em_om:*           85:  gb_pr1:*
         43:  em_or:*           86:  gb_pr2:*
                                87:  gb_pr3:*
                                88:  gb_pr4:*
                                89:  gb_pr5:*
                                90:  gb_pr6:*
                                91:  gb_pr7:*
                                92:  gb_pr8:*
                                93:  gb_pr9:*
                                94:  gb_ro1:*
                                95:  gb_ro2:*
                                96:  gb_in4:*
                                97:  gb_pr10:*
                                98:  em_ba3:*


                                SUMMARIES

                                                              %
Result                          Query  Match  Length  DB  ID              Description
 No.      Score  Match                                   ----------              -----------

No matches found

Search completed: N_____er  7, 2001, 04:50:11
Job time: 4906 se_

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          November  7, 2001, 03:48:16 ; Search time 129.24 Seconds
                                                (without alignments)
                                                2259.163 Million cell updates/sec

Title:           US-09-521-640-2
Perfect score:   465
Sequence:        1 ggggagtatcgtgagttcac..........cccttcttgttcaaaacacn 465

Scoring table:   OLIGO_NUC
                 Gapop 60.0 , Gapext 60.0

Se...ned:        730101 seqs, 31395080 9 residues

Word size :      30

Total number of hits satisfying chosen parameters:      0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :       N_Geneseq_0601:*
                 1:  /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
                 2:  /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
                 3:  /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
                 4:  /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
                 5:  /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
                 6:  /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
                 7:  /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
                 8:  /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
                 9:  /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
                10:  /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
                11:  /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
                12:  /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
                13:  /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
                14:  /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
                15:  /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
                16:  /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
                17:  /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
                18:  /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
                19:  /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
                20:  /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
                21:  /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
                22:  /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES

Result       Query
No.    Score Match Length DB  ID                          Description
------ ----- ----- ------ --  --                          -----------

No matches found

Search completed: November  7, 2001, 04:52:32
Job time: 3856 sec

```
                  GenCore version 4.5
             Copyright (c) 1993 - 2000  Compugen Ltd.


OM nucleic - nucleic search, using sw model

Run on:       November  7, 2001, 03:27:20 ; Search time 1162.82 seconds
                                            (without alignments)
                                            3780.101 Million cell updates/sec

Title:        US-09-521-640-2
Perfect score: 465
Sequence:      1 ggggagtatcgtgagttcac........cccttcttgttcaaaacacn 465

Scoring table: OLIGO_NUC
               Gapop 60.0 , Gapext 60.0

Seq ned:      1028115 seqs, 4726426750 residues

Word size :   30

Total number of hits satisfying chosen parameters:       0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :    EST:*

     1: gb_est1:*            44: em_esthum10:*
     2: gb_est2:*            45: em_esthum11:*
     3: gb_est3:*            46: em_esthum12:*
     4: gb_est4:*            47: em_esthum13:*
     5: gb_est5:*            48: em_esthum14:*
     6: gb_est6:*            49: em_esthum15:*
     7: gb_est7:*            50: em_esthum16:*
     8: gb_est8:*            51: em_esthum17:*
     9: gb_est9:*            52: em_esthum18:*
    10: gb_est10:*           53: em_esthum19:*
    11: gb_est11:*           54: em_esthum20:*
    12: gb_est12:*           55: em_esthum21:*
    13: gb_est13:*           56: em_esthum22:*
    14: gb_est14:*           57: em_esthum23:*
    15: gb_est15:*           58: em_esthum24:*
    16: gb_est16:*           59: em_esthum25:*
    17: gb_est17:*           60: em_esthum26:*
    18: gb_est18:*           61: em_esthum27:*
    19: gb_est19:*           62: em_esthum28:*
    20: gb_est20:*           63: em_estin1:*
    21: gb_est21:*           64: em_estin2:*
    22: gb_est22:*           65: em_estin3:*
    23: gb_est23:*           66: em_estin4:*
    24: gb_est24:*           67: em_estin5:*
    25: gb_est33:*           68: em_estom1:*
    26: gb_est34:*           69: em_estom2:*
    27: gb_est35:*           70: em_estov1:*
    28: gb_est36:*           71: em_estov2:*
    29: gb_est37:*           72: em_estpl1:*
    30: gb_est38:*           73: em_estpl2:*
    31: gb_est39:*           74: em_estpl3:*
    32: gb_est40:*           75: em_estpl4:*
    33: em_estba:*           76: em_estpl5:*
    34: em_estfun:*          77: em_estpl6:*
    35: em_esthum1:*         78: em_estpl7:*
    36: em_esthum2:*         79: em_estpl8:*
    37: em_esthum3:*         80: em_estpl9:*
    38: em_esthum4:*         81: em_estpl10:*
    39: em_esthum5:*         82: em_estro1:*
    40: em_esthum6:*         83: em_estro2:*
    41: em_esthum7:*         84: em_estro3:*
    42: em_esthum8:*         85: em_estro4:*
    43: em_esthum9:*         86: em_estro5:*
                             87: em_estro6:*
                             88: em_estro7:*
                             89: em_estro8:*
                             90: em_estro9:*
                             91: em_estro10:*
                             92: em_estro11:*
                             93: em_estro12:*
                             94: em_estro13:*
                             95: em_estro14:*
                             96: em_estro15:*
                             97: em_estro16:*
                             98: em_estro17:*
                             99: em_estro18:*
                            100: em_estro19:*
                            101: em_estro20:*
                            102: gb_est25:*
                            103: gb_est26:*
                            104: gb_est27:*
                            105: gb_est28:*
                            106: gb_est29:*
                            107: gb_est30:*
                            108: gb_est31:*
                            109: gb_est32:*
                            110: gb_est41:*
                            111: gb_est42:*
                            112: gb_est43:*
                            113: gb_est44:*
                            114: gb_est45:*
                            115: gb_est46:*
                            116: gb_est47:*
```

117:  gb_est48:*
118:  gb_est49:*
119:  gb_est50:*
120:  gb_est51:*
121:  gb_est52:*
122:  gb_est53:*
123:  gb_est54:*
124:  gb_est55:*
125:  gb_est56:*
126:  gb_est57:*
127:  gb_est58:*
128:  gb_est59:*
129:  gb_est60:*
130:  gb_est61:*
131:  gb_est62:*
132:  gb_est63:*
133:  gb_est64:*
134:  gb_est65:*
135:  gb_est66:*
136:  gb_est75:*
137:  gb_est76:*
138:  gb_est77:*
139:  gb_est78:*
140:  gb_est79:*
141:  gb_est80:*
142:  gb_est81:*
143:  gb_est82:*
144:  gb_est83:*
145:  gb_est84:*
146:  gb_est85:*
147:  gb_est86:*
148:  gb_est87:*
149:  gb_est88:*
150:  gb_est89:*
151:  gb_est90:*
152:  gb_est99:*
153:  gb_est100:**
154:  gb_est101:**
155:  gb_est102:**
156:  gb_est103:**
157:  gb_est104:**
158:  gb_est105:**
159:  gb_est106:**
160:  gb_est67:*
161:  gb_est68:*
162:  gb_est69:*
163:  gb_est70:*
164:  gb_est71:*
165:  gb_est72:*
166:  gb_est73:*
167:  gb_est74:*
168:  gb_est91:*
169:  gb_est92:*
170:  gb_est93:*
171:  gb_est94:*
172:  gb_est95:*
173:  gb_est96:*
174:  gb_est97:*
175:  gb_est98:*
176:  em_esthum29:*
177:  em_esthum30:**
178:  em_esthum31:**
179:  em_esthum32:**
180:  em_esthum33:**
181:  em_estom3:*
182:  em_estpl11:*
183:  em_estro21:**
184:  em_estro22:**
185:  em_estro23:**
186:  em_htc:*
187:  gb_est107:*
188:  gb_est108:**
189:  gb_est109:**

190:  gb_est110:*
191:  gb_est111:*
192:  gb_htc:*
193:  em_gss_fun:*
194:  em_gss_hum1:*
195:  em_gss_hum2:*
196:  em_gss_hum3:*
197:  em_gss_hum4:*
198:  em_gss_hum5:*
199:  em_gss_hum6:*
200:  em_gss_hum7:*
201:  em_gss_hum8:*
202:  em_gss_hum9:*
203:  em_gss_inv1:*
204:  em_gss_inv2:*
205:  em_gss_inv3:*
206:  em_gss_other:*
207:  em_gss_pln1:*
208:  em_gss_pln2:*
209:  em_gss_pro:*
210:  em_gss_rod1:*
211:  em_gss_rod2:*
212:  em_gss_rod3:*
213:  em_gss_rod4:*
214:  em_gss_rod5:*
215:  em_gss_vrt1:*
216:  em_gss_vrt2:*
217:  em_gss_vrt3:*
218:  gb_gss1:*
219:  gb_gss2:*
220:  gb_gss3:*
221:  gb_gss4:*
222:  gb_gss5:*
223:  gb_gss6:*
224:  gb_gss7:*
225:  gb_gss8:*
226:  gb_gss9:*
227:  gb_gss10:*
228:  gb_gss11:*
229:  gb_gss12:*
230:  gb_gss13:*
231:  gb_gss14:*
232:  gb_gss15:*
233:  gb_gss16:*
234:  gb_gss17:*
235:  gb_gss18:*
236:  gb_gss19:*
237:  gb_gss20:*
238:  gb_gss21:*
239:  gb_gss22:*
240:  gb_gss23:*
241:  gb_gss24:*
242:  gb_gss25:*
243:  gb_gss26:*
244:  gb_gss27:*
245:  gb_gss28:*
246:  gb_gss29:*
247:  gb_gss30:*
248:  gb_gss31:*
249:  gb_gss32:*
250:  gb_gss33:*
251:  gb_gss34:*
252:  em_gss_inv4:*
253:  em_gss_rod6:**
254:  em_gss_rod7:*
255:  em_gss_rod8:**
256:  gb_gss35:*
257:  gb_gss36:*
258:  gb_gss37:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|-----|-----|-------------|

No matches found

Search completed: November 7, 2001, 04:30:08
Job time: 3768 sec

```
             GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      November  7, 2001, 03:24:30 ; Search time 1194.37 Seconds
                                            (without alignments)
                                            6022.006 Million cell updates/sec

Title:          US-09-521-640-2
Perfect score:  465
Sequence:       1 ggggagtatcgtgagttcac........cccttcttgttcaaaacacn 465

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Sea..hed:       1344157 seqs, 773387458B residues

Total number of hits satisfying chosen parameters:     2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
              1: gb_ba1:*
              2: gb_ba2:*
              3: gb_ba3:*
              4: gb_in1:*
              5: gb_in2:*
              6: gb_in3:*
              7: gb_om:*
              8: gb_ov:*
              9: gb_pat1:*
             10: gb_pat2:*
             11: gb_ph:*
             12: gb_pl1:*
             13: gb_pl2:*
             14: gb_pl3:*
             15: gb_pl4:*
             16: em_ba1:*
             17: em_ba2:*
             18: em_fun:*
             19: em_htgo_hum:*
             20: em_htgo_inv:*
             21: em_htgo_rod:*
             22: em_htg_hum1:*
             23: em_htg_hum2:*
             24: em_htg_hum3:*
             25: em_htg_hum4:*
             26: em_htg_hum5:*
             27: em_htg_hum6:*
             28: em_htg_hum7:*
             29: em_htg_hum8:*
             30: em_htg_inv1:*
             31: em_htg_inv2:*
             32: em_htg_other:*
             33: em_htg_rod:*
             34: em_hum1:*
             35: em_hum2:*
             36: em_hum3:*
             37: em_hum4:*
             38: em_hum5:*
             39: em_hum6:*
             40: em_hum7:*
             41: em_in:*
             42: em_om:*
             43: em_or:*
             44: em_ov:*
             45: em_pat:*
             46: em_ph:*
             47: em_pl:*
             48: em_ro:*
             49: em_sts:*
             50: em_sy:*
             51: em_un:*
             52: em_vi:*
             53: gb_sts1:*
             54: gb_sts2:*
             55: gb_sts3:*
             56: gb_sy:*
             57: gb_un:*
             58: gb_vi1:*
             59: gb_vi2:*
             60: gb_htg1:*
             61: gb_htg2:*
             62: gb_htg3:*
             63: gb_htg4:*
             64: gb_htg5:*
             65: gb_htg6:*
             66: gb_htg7:*
             67: gb_htg8:*
             68: gb_htg9:*
             69: gb_htg10:*
             70: gb_htg11:*
             71: gb_htg12:*
             72: gb_htg13:*
             73: gb_htg14:*
             74: gb_htg15:*
             75: gb_htg16:*
             76: gb_htg17:*
             77: gb_htg18:*
             78: gb_htg19:*
             79: gb_htg20:*
             80: gb_htg21:*
             81: gb_htg22:*
             82: gb_htg23:*
             83: gb_htg24:*
             84: gb_htg25:*
             85: gb_pr1:*
             86: gb_pr2:*
             87: gb_pr3:*
             88: gb_pr4:*
             89: gb_pr5:*
             90: gb_pr6:*
             91: gb_pr7:*
             92: gb_pr8:*
             93: gb_pr9:*
             94: gb_ro1:*
             95: gb_ro2:*
             96: gb_in4:*
             97: gb_pr10:*
             98: em_ba3:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Mat | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
|   1 | 54.8 | 11 | 1/1919 | 65 | AC018552 | AC018552 Homo sapi |
|   2 | 52.6 | 11 | 7218 | 10 | I66494 | I66494 Sequence 14 |
| c  3 | 37.2 | 8 | 143786 | 6 | CELY71H2AM | AC024859 Caenorhab |
| c  4 | 37.2 | 8 | 48960 | 60 | AC006896 | AC006896 Caenorhab |
| c  5 | 36.8 | 7 | 485 | 8 | AF309412 | AF309412 Oncorhync |
| c  6 | 36.4 | 7 | 215283 | 52 | AC011966 | AC011966 Homo sapi |
| c  7 | 36.2 | 7 | 40552 | 4 | AC084469 | AC084469 Caenorhab |
| c  8 | 36 | 7 | 174296 | 63 | AC013805 | AC013805 Homo sapi |

```
  9    36    7.7  334796  70  AC026673       AC026673  Homo sapi
 10  35.8    7.7  132449  90  AL365272       AL365272  Human DNA
 11  35.8    7.7  171350  72  AC058816       AC058816  Homo sapi
 12  35.8    7.7  172307  71  AC044842       AC044842  Homo sapi
c13  35.8    7.7  182529  82  AL589962       AL589962  Homo sapi
c14  35.2    7.6  180538  88  AC073897       AC073897  Homo sapi
 15    35    7.5  160069  83  AP001977       AP001977  Homo sapi
 16    35    7.5  160800  83  AP002011       AP002011  Homo sapi
 17    35    7.5  183674  64  AC016996       AC016996  Homo sapi
 18    35    7.5  189078  83  AP001982       AP001982  Homo sapi
 19  34.8    7.5   16121   6  CEF40F11       273426    Caenorhabdi
 20  34.8    7.5   85953  83  CEH04L24       Z97189    Caenorhabdi
c21  34.8    7.5  107244  83  CEY24F12       AL022277  Caenorhab
c22  34.8    7.5  131318  83  CEY42E10       Z99944    Caenorhabdi
c23  34.8    7.5  169689  72  AC051616       AC051616  Mus muscu
 24  34.6    7.4    1246  89  AF244570       AF244570  Homo sapi
 25  34.6    7.4    1480  89  AF244571       AF244571  Homo sapi
c26  34.6    7.4    1745  13  AF168859       AF168859  Mayaca au
c27  34.6    7.4  186647  69  AC025573       AC025573  Homo sapi
 28  34.4    7.4   32488   6  CELT12B5       CELT12B5  Caenorhab
c29  34.4    7.4   79531  90  AL390037       AL390037  Human DNA
c30  34.2    7.4   73006  62  AC011388       AC011388  Homo sapi
  1  34.2    7.4  104271  60  AC008472       AC008472  Homo sapi
  2  34.2    7.4  129389  92  HS526I14       Z82214    Human DNA s
c 3  34.2    7.4  209336  74  AC073116       AC073116  Homo sapi
 34  34.2    7.4  298575  84  HSBA22B4       AL080244  Homo sapi
c35  33.8    7.3    1933   6  FPA224788      AJ224788  Flabellin
c36  33.8    7.3    1941   5  ESA224786      AJ224786  Eubranchu
c37  33.8    7.3    1969   6  AF249198       AF249198  Flabellin
c38  33.8    7.3    2087   6  MLA224784      AJ224784  Melibe le
c39  33.8    7.2  164399  74  AC073321       AC073321  Homo sapi
c40  33.6    7.2    7764  89  AF225896       AF225896  Homo sapi
c41  33.6    7.2   17245   7  OCAJ1588       AJ001588  Oryctolag
 42  33.6    7.2   62043  90  AL161893       AL161893  Human DNA
c43  33.6    7.2  138039  64  AC016786       AC016786  Homo sapi
c44  33.6    7.2  152696  69  AC025195       AC025195  Homo sapi
c45  33.6    7.2  153279  65  AC018395       AC018395  Homo sapi
```

ALIGNMENTS

```
RESULT    1
AC018552
LOCUS       AC018552     171919 bp    DNA              HTG       26-SEP-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11-405F3, WORKING DRAFT
            SEQUENCE, 19 unordered pieces.
ACCESSION   AC018552
VERSION     AC018552.4  GI:10305141
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171919)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Human Chromosome 16
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 171919)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (14-DEC-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 26, 2000 this sequence version replaced gi:8576099.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 0
            Center clone name: RPCI-11-405F3
            -----
```

```
            Summary Statistics
            Consensus quality: 130458 bases at least Q40
            Consensus quality: 142777 bases at least Q30
            Consensus quality: 149702 bases at least Q20
            Estimated insert size: 171000; agarose-fp estimation
            Estimated insert size: 170119; sum-of-contigs estimation
            Quality coverage:      4.73 in Q20 bases; agarose-fp estimation
            Quality coverage:      4.76 in Q20 bases; sum-of-contigs estimation.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 19 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *          1     1194: contig of 1194 bp in length
          *       1195     1294: gap of unknown length
          *       1295     2441: contig of 1147 bp in length
          *       2442     2541: gap of unknown length
          *       2542     3628: contig of 1087 bp in length
          *       3629     3728: gap of unknown length
          *       3729     4822: contig of 1094 bp in length
          *       4823     4922: gap of unknown length
          *       4923     6068: contig of 1146 bp in length
          *       6069     6168: gap of unknown length
          *       6169     7451: contig of 1283 bp in length
          *       7452     7551: gap of unknown length
          *       7552     8563: contig of 1012 bp in length
          *       8564     8663: gap of unknown length
          *       8664    10826: contig of 2163 bp in length
          *      10827    10926: gap of unknown length
          *      10927    12292: contig of 1366 bp in length
          *      12293    12392: gap of unknown length
          *      12393    13481: contig of 1089 bp in length
          *      13482    13581: gap of unknown length
          *      13582    14780: contig of 1199 bp in length
          *      14781    14880: gap of unknown length
          *      14881    18391: contig of 3511 bp in length
          *      18392    18491: gap of unknown length
          *      18492    21416: contig of 2925 bp in length
          *      21417    21516: gap of unknown length
          *      21517    29200: contig of 7684 bp in length
          *      29201    29300: gap of unknown length
          *      29301    43025: contig of 13725 bp in length
          *      43026    43125: gap of unknown length
          *      43126    55543: contig of 12418 bp in length
          *      55544    55643: gap of unknown length
          *      55644    70587: contig of 14944 bp in length
          *      70588    70687: gap of unknown length
          *      70688    94927: contig of 24240 bp in length
          *      94928    95027: gap of unknown length
          *      95028   171919: contig of 76892 bp in length.
FEATURES             Location/Qualifiers
     source          1..171919
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-405F3"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT     40913 a   42871 c   43420 g   42804 t   1911 others
ORIGIN
```

```
Query Match          11.8%;    Score 54.8;   DB 65;    Length 171919;
Best Local Similarity 61.0%;   Pred. No. 1.3e-05;
Matches   89;  Conservative   0;  Mismatches  57;  Indels   0;  Gaps   0;

Qy    113  gagaaactgcgtacccgctaatctcttcacacatcccctatcgcactggcggtatctca 172
               || |||||||||| |||||   ||||| |     |  ||||   |  |
Db  10105  GAAAACCCTGCGTACCAACTAATTGCCTGGGGACTTCCCTTTACCACTTCGGTAAGGCAA 10164

Qy    173  atgaccaccatcgcgtccaaccatgccaactaatgcaatggccctgatcggattctc 232
```

```
Db   10165  AAGGCCAACGATCCCTTCCACACATGCCACCTGAAGGCAATGGCCCAAACGGATTTCTC 10224
                  | | |||   |||||   ||||||| ||||||| || ||||||||||| |||| ||  | |||||||
Qy     233  ttaccattgtcggattacccgctat 258
                  ||| ||||  ||||||||||||||| ||
Db   10225  TTTCACATTGCGGGATTTCACCAGAT 10250
```

RESULT   2

LOCUS        I66494              7218 bp      DNA              PAT       28-DEC-1997
DEFINITION   Sequence 14 from patent US 5670367.
ACCESSION    I66494
VERSION      I66494.1   GI:2724471
KEYWORDS     .
SOURCE       Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE    1  (bases 1 to 7218)
  AUTHORS    Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE      Recombinant fowlpox virus
  JOURNAL    Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1..7218
                     /organism="unknown"

BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN

```
Query Match            11.3%;  Score 52.6;  DB 10;  Length 7218;
Best Local Similarity   6.5%;  Pred. No.4.8e-05;
Matches     28;  Conservative 220;  Mismatches 180;  Indels  0;  Gaps  0;

Qy     37  agactcgctgtaagacccggtccagccgatcccctttatgagcggctaagtcactgccg  96
               ::: :: :::: : :::::: ::: :::  | ||| ::::::::::::::::::::::
Db   1036  AGCTTCGGCTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYY 1095

Qy     97  cgcttacacgcggactgagaaacctgcgtaccgcgtaatctcttcacacatcccctatcg 156
               :::::: ::::::::::::::::YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
Db   1096  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1155

Qy    157  cactgcggtatctcaatgaccaccatcgccttccaaccatcggccaactaatggcaatggc 216
               ::: ::: :: :::::::::::::::::::::::: :::::::::::::::::::::::
Db   1156  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1215

Qy    217  cctgatcggattctcttaccattgtcggattacccgctcatggcgcttcatcaacagag 276
               ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db   1216  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1275

Qy    277  gatgccctagtatcacctaccacaccccgtgacaaccttggccttaaataaaacct 336
               :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db   1276  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1335

Qy    337  atctttgtcatttcggaataaacttcctgtcttcagcaaatgaagaaccccctgctgag 396
               :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db   1336  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1395

Qy    397  tcctttgggcgggccctttgcacgctttggacttgttatcaaaaaaaccctctgtt 456
               :::: :: ::::::::::::::::::::::::::::::::: ::: |  |||| ::|
Db   1396  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAATTCTTCTATCTCT 1455

Qy    457  caaaacac 464
               ||| ||
Db   1456  TTAACTAC 1463
```

RESULT   3

CELY71H2AM/C

LOCUS        CELY71H2AM    143786 bp      DNA              INV       06-APR-2001
DEFINITION   Caenorhabditis elegans cosmid Y71H2AM, complete sequence.
ACCESSION    AC024859
VERSION      AC024859.1   GI:7140422
KEYWORDS     HTG.
SOURCE       Caenorhabditis elegans.
  ORGANISM   Caenorhabditis elegans
             Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
             Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE    1  (bases 1 to 143786)
  AUTHORS    The C. elegans Genome Sequencing Consortium, Washington University
             Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,
             Hinxton, U.K., C.
  TITLE      Genome sequence of the nematode C. elegans: a platform for
             investigating biology. The C. elegans Sequencing Consortium
  JOURNAL    Science 282 (5396), 2012-2018 (1998)
  MEDLINE    99069613
REFERENCE    2  (bases 1 to 143786)
  AUTHORS    Waterston.R.H.
  TITLE      Direct Submission
  JOURNAL    Submitted (01-MAR-2000) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
REFERENCE    3  (bases 1 to 143786)
  AUTHORS    Waterston.R.
  TITLE      Direct Submission
  JOURNAL    Submitted (06-APR-2001) Department of Genetics, Washington
             University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT      Submitted by:
             Genome Sequencing Center
             Department of Genetics, Washington University,
             St. Louis, MO 63110, USA, and
             Sanger Centre, Hinxton Hall
             Cambridge CB10 1RQ, England
             e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

             NOTICE: This sequence may not be the entire insert of this clone.
             It may be shorter because we only sequence overlapping sections
             once, or longer because we provide a small overlap between
             neighboring submissions.

             NOTES:

             Coding sequences below are predicted from computer analysis, using
             the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES             Location/Qualifiers
     source          1..143786
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="CELY71H2AM"
     gene            complement(2225..8816)
                     /gene="Y71H2AM.9"
     CDS             complement(join(2225..2326,2429,..2493,3402,..3513,
                     3572..3696,5645..5789,6811..7111,7191..7389,7443..7489,
                     8689..8816))
                     /gene="Y71H2AM.9"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="hypothetical protein Y71H2AM.9"
                     /protein_id="AAK29970.1"
                     /db_xref="GI:13559766"
                     /translation="MKSLMKMVNFEGFFKGFASEFLAKKLVKNSKKKIKIKAIIWAGS
                     KMIPMSLVISNVNTVFRRKNPILMVSNVRCLSSNQPTDDDMHISKRKLKKLFKVAQLS
                     KKSRSYENDKIDKLSAMAEFGLKESDLEDLPSHSKPVLSLTHHGKSNHQLYTLTEVYH
                     LALKVHGSTEALEAHRPLVAMRKELTDAERMRMRVETTNQHKGADRVVAIAFTLN
                     CDMAMKFTAAYLTGSKSLFAEAIHSAMDTCNQLILLLGIRYSAKNPDLLFPVGYGNM
                     YVTSLISGCGIMAFGCGLSMYHGISGLLHPEPLEPLTYAYYALFMSLCFQGTSAITA
                     LFVFAKKAKNISIWNYVRTSADPSLNVVLLEDTAAVTGVGIALCAISLSSFLNSPI
                     LDGSIVIGALLGS"
     gene            ..14531
                     /gene="Y71H2AM.8"
     CDS             join(9613..9658,9753..9909,10763..10879,12629..13172,
                     14056,.14418,14490..14531)
                     /gene="Y71H2AM.8"
```

```
          /codon_start=1
          /evidence=not_experimental
          /product="Hypothetical protein Y71H2AM.8"
          /protein_id="AAK29971.1"
          /db_xref="GI:13559767"
          /translation="MAPSPTAELEDQILGKIATLFEETLSKQPPSNIVELMKEHVEIR
          RILNDVISIQAKLGPTDERLANSKRDADSIKTFLAWADGVGIARNNVTIGSTKTAGLS
          LQATGPIPKSHIVARVPRHAMITLDLAKKSSLLKKAFEKDPIVGGMDNVGLALFLACQ
          WIQNEKSWKSYISILPTTFPTPLFYSEEQLQLKPSPIFEEAILFYRTISRQFCYFL
          LAIAKNKIYBAAQRRKDARNAMETPIFYNVPFNVANFTPKLYFWAVGVVTTRVNMVPS
          ENQVGEDGNGVIIPALIPVLDMANHENVLTDVLTEPIEDLVCYSPEEECAVITSHCDV
          KAGNEVTIFYGCRSKGEHLLHNGFVPIYHGKFDVLKLKIGIPKTDKTLDAKKLLIQKF
          VKKVYCAGNIFHVDLYNCKNMRISLKIRFC"
gene      complement(16532..23598)
          /gene="Y71H2AM.10"
CDS       complement(join(16532..16726,17546..17679,18063..18187,
          18239..18398,19350..19840,20512..21131,21447..21702,
          21749..21883,22158..22268,22316..22555,23146..23251,
          23491..23598))
          /gene="Y71H2AM.10"
          /note="coded for by C. elegans cDNA yk285g2.5; coded for
          by C. elegans cDNA yk169a12.5; coded for by C. elegans
          cDNA yk285g2.3; coded for by C. elegans cDNA yk169a12.3"
          /codon_start=1
          /product="Hypothetical protein Y71H2AM.10"
          /protein_id="AAK29969.1"
          /db_xref="GI:13559765"
          /translation="MATVYSADLSSSSRRVFEEIPLHCAVRVRIITEKVVTRPYVDLL
          AVVDESLSRHVGAICVHAQITSSTSAIELGKCYLNHEHDQKSSCVIRIPVFSWFPVD
          QNRTSVLSVSVTVSEKCDQNFHDLPQHLIEVNSRIAKQKIDWLANATETSVTLLSTAS
          QAFSQNSMQTLFLHVKSWANETQPMEIRLWVDSRMSIETVYPTSSNWTIRVSSASRPF
          FYTSLVCTPKERMTGFNDNIVAILIKMVSSTDAIKDDVILHWHVIIGPKSSEPPPDDH
          KVATKFSVIADEVAAVVIVPKVTCGRSQYITFHAQKKTTSHRKELMNLAVISGIGVT
          SSLRIFTISIGAKAEDVTTQSHCISSDANIIKVSPTCSSVYLDGSESNGSSDAQYYAH
          YLRYTTYSFRVWFPKLPLKIWMSSSTLSIKNWKVGFWRDLPLGGGVKRSRAARQFA
          CVNRFQHSHVVLASLWIEDKTGDQLYLSSHKSILFDVTNIVHNTLQISNRTVANVK
          FYEGRAKVIGENVGLAKLIVRNAKKSMDLVSENISVQNKEVSTTGLSARPICDTSFRI
          LPILFSPAFFKVEIAHSKSLTKLYQQCSIFASVSYSDATWEPLNDLDSSYFEMSAHSD
          NERALAVSHHASKVHVIAIDENWPLPVVEISLQSSAQCASTINGASPAALAVTVLNVP
          IKINNSFPSATDLDSFSTTIAPSDSLPSIPFHIFVLTIIGLIIILFISFVRRSAAF
          KGYEQLVPFFSRLSSSSGSNSRQEETNEWWLSGPQPPSSTISSGYSGNKSTAERQS
          SNGDDPSRTSISYHGSEISVFIAPSQGNVVVNQSSRHPRYTLVDSNSDHNLARIVPKE
          DRWTTGGHEQFHTWTMKQRGGGGGRMEAPIRESIA"
gene      complement(27014..33736)
          /gene="Y71H2AM.11"
CDS       complement(join(27014..27225,27780..27833,27889..28391,
          29511..30243,32146..32213,33337..33442,33535..33658,
          33711..33736))
          /gene="Y71H2AM.11"
          /note="coded for by C. elegans cDNA yk646h6.5; coded for
          by C. elegans cDNA yk646h6.3"
          /codon_start=1
          /product="Hypothetical protein Y71H2AM.11"
          /protein_id="AAK29983.1"
          /db_xref="GI:13559779"
          /translation="MBQIDAVFKSIDERQDEFIDLLRESVAIQSVSADPARRGDCVRM
          SEWARDQLKTLGVETSLWELGQQTLPSGEQLPLPPAVFGVYGRDKSKKTLLIYGHDV
          QPAEKEDGWNTNPFELTEIDGKLFGRGSTDDKGPVIAWIAVLKVLQTLGIDLPINIKF
          VLECMESSSEGLDKGLEDNIDKISDVTFSCISDNYWLGRNKCLTPEALNDLMWVMS
          QLVTVDGQILIPGIAELVAPLTKDEDIYERKIDFCVDTFKNETGSHGLMSDNKNLLM
          NRWRYPSLSLHGVEGAFSQPGAKTVIPAKVVGKFSLRIVPHMTPEATDKLVNSYLDSL
          WAKRKSPNTFKVTSGHGGMPWVADFRDANFSAGSRAIERDLENFRKHFSKKLNFPPKL
          YGMTPDFTREGGSIPVTLTIQDLTKSPVMLLPIGASDDMAHSQNEKINRDNFVKGMKV
          LAAYLFELAAA"
gene      35242..40205
          /gene="Y71H2AM.7"
CDS       join(35242..35313,35416..35535,37225..37334,38048..38494,
          39905..40097,40170..40205))
          /gene="Y71H2AM.7"
          /codon_start=1
          /evidence=not_experimental
          /product="Hypothetical protein Y71H2AM.7"
          /protein_id="AAK29972.1"
```

```
          /db_xref="GI:13559768"
          /translation="MAVDIRGVYIGNSMFILQKFVNFKLIPNSERQNAANQSDFDFM
          KPKLVEYVFTVCVRLRLPNESLSSROVNTLRKIGLPYTISAVLESEORVFKLIGFKM
          PDSPLDACEMALKVLFTMKKRGMIDEEKYNDLWQHTLIVLDVCFINHIELVERFIRK
          CPAICRTEERHGFSLQLFPHGGRGKVVSRPWQRGRQVLTTIYLALFSACFLSFFIDFF
          RFFLIKLNKYFLQMLKQFPSKKIMYSVGKQRLNISRFKWDIOLLAATVQTAYILLL
          GTSQIANVSVIINNLRCDNAVVEPLKQSIIELACAKKNESIPECSTSS"
gene      complement(40206..42546)
          /gene="Y71H2AM.12"
CDS       complement(join(40206..40261,41089..41240,41746..41854,
          41904..42012,42332..42444,42498..42546))
          /gene="Y71H2AM.12"
          /note="coded for by C. elegans cDNA yk301b4.5; coded for
          by C. elegans cDNA yk234f4.5; coded for by C. elegans cDNA
          yk301b4.3"
          /codon_start=1
          /product="Hypothetical protein Y71H2AM.12"
          /protein_id="AAK29982.1"
          /db_xref="GI:13559778"
          /translation="MDTIKTVKVVVVGESGAGKTALLTCFLDNTFEFDPLTTIGIDFK
          HKIVQLNDGQSIRLQLWDTAGQERFRQLAPAYIRSARVALLVIDLSDENCVEHIIRWK
          GIIDKNKSDFTSTIIVGNKHDLVSEKRSPRLTAIIRETNDEYIETSAKMRKNIKKLFS
          SVACRPPEHETSQIILLNEPRPVESATKRCCQRW"
gene      complement(42822..48202)
          /gene="Y71H2AM.13"
CDS       complement(join(42822..43008,43541..43770,44200..44526,
          44910..45066,45113..45297,46544..46943,47007..47099,
          47964..48023,48089..48202))
          /gene="Y71H2AM.13"
          /note="coded for by C. elegans cDNA yk75h9.5; coded for by
          C. elegans cDNA yk647e3.5; coded for by C. elegans cDNA
          yk253h4.5; coded for by C. elegans cDNA yk513a6.5; coded
          for by C. elegans cDNA yk100g10.5; coded for by C. elegans
          cDNA yk65c4.5; coded for by C. elegans cDNA yk169d11.5;
          coded for by C. elegans cDNA yk513a6.3; coded for by C.
          elegans cDNA yk169d11.3; coded for by C. elegans cDNA
          yk253h4.3; coded for by C. elegans cDNA yk647e3.3; coded
          for by C. elegans cDNA yk75h9.3; coded for by C. elegans
          cDNA yk378b12.3"
          /codon_start=1
          /product="Hypothetical protein Y71H2AM.13"
          /protein_id="AAK29968.1"
          /db_xref="GI:13559764"
          /translation="MGGTVSKHTTPHYPPSROVLTSYGPIEGRRLIHEGEKQVDAFQG
          IPYAAPIGNLRFALPQPHEKWTEVRETKSFGARGIQKDHVLSPKTSPQSEDNLTLNI
          FFPVWTPKNETGPVILYIHGGFVSDSAHKYGDMSICQHLVTKDVVVVTIQYRLGFL
          GFWTTGDSSIPDNVALHDVFALKWVKENIGLFNGDPNNITLMGQSAGGASVDFLSIS
          PVSRDLFQKVIPMGGNASCSWAIHPRPLNACRNRAQEIGVFDGMNTLDWVEKLRELPA
          DKFASALNMEAVDTKTDPELLIGPKYDNLFIPRPVMELRKEAPMKPRLLGCAKSEGLV"
```

```
Query Match     8.0%;      Score 37.2;      DB 6;      Length 143786;
Best Local Similarity    53.1%;      Pred. No. 2.8;
Matches    78;      Conservative    0;      Mismatches    0;      Indels    69;      Gaps    0;

Qy    317  ttggccttaaataaacctatctttgtcattcggaataaacttcctgtctttcagca 376
                ||||| || |    | | | |||  |  ||||  || |   |||||| ||| |||
Db  79188  TTTTCCAAAAAAAAACCCAATTTTGAAATTTGGAAAAATATTCCATTTTTTAAACA 79129

Qy    377  aatgaagaaccccgctcngtcctttgggcgggccttttgcacgctttttgactgtta 436
                || | |  | ||| ||  |  |    |||  |     | | | |||||||  | |||
Db  79128  AATTTCAAAAAAATCATTTTTTGTTGAAATTTTGGAAACTATTTTTCGATTTTTTT 79069

Qy    437  tcaaaaaaccctcttgttcaaaca 463
                |||||||  |   ||| |||||||
Db  79068  GAAAAAAATCAATTATTCTGAAAAA 79042


RESULT    4
AC006896/c
LOCUS       AC006896    298960 bp    DNA          HTG          26-FEB-1999
DEFINITION  Caenorhabditis elegans clone Y71H2X, *** SEQUENCING IN PROGRESS
            ***, 15 unordered pieces.
ACCESSION   AC006896
```

```
Db 153784  AATTCAAAAAAAATCATTTTTTTTGTGAAATTTTTGGAAACTATTTTCGATTTTTT 153725
                 |||||||||||| ||||||||| | | ||||  | | | |||||
Qy     437  tcaaaaaaccctctgttcaaaaca 463
                 |||||||| ||||||||| | |||||
Db 153724  GAAAAAAATCAATTATTCTGAAAAA 153698

RESULT   5
AF309412/c
LOCUS       AF309412        485 bp    DNA                  VRT        24-OCT-2000
DEFINITION  Oncorhynchus mykiss 18S ribosomal RNA gene, partial sequence.
ACCESSION   AF309412
VERSION     AF309412.1  GI:10954038
KEYWORDS    .
SOURCE      rainbow trout.
  ORGANISM  Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1  (bases 1 to 485)
  AUTHORS   Jones,I., Kille,P., Wigham,T. and Sweeney,G.E.
  TITLE     Cloning and characterization of rainbow trout 18s rRNA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 485)
  AUTHORS   Jones,I., Kille,P., Wigham,T. and Sweeney,G.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2000) Department of Biosciences, Cardiff
            University, Museum Avenue, Cardiff CF1 3US, UK
FEATURES             Location/Qualifiers
     source          1..485
                     /organism="Oncorhynchus mykiss"
                     /db_xref="taxon:8022"
                     <1..>485
     rRNA            /product="18S ribosomal RNA"
BASE COUNT      119 a    117 c    144 g    105 t
ORIGIN

Query Match                7.9%;    Score 36.8;  DB 8;   Length 485;
Best Local Similarity     53.5%;    Pred. No. 2.6;
Matches   99;  Conservative   0;  Mismatches   83;  Indels   3;  Gaps   1;

Qy    229  tctcctaccattgtcggattacccgctatggcgcttcatcaacagaggatgcccctagta 288
                ||||||||||  | | ||| |  | ||| |||||  || |||||| || ||| ||||| |||
Db    364  TCTCGTTCGTTATCGGAATTAACCAGACAAAATGCTCCACCAACTAAGAACGGCCCATGCA 305

Qy    289  tcacccaccaccacccgtgaccaacctggcctttaaataaacctatctttgtcatt 348
                || | ||| | |||||| | | |||| |||| || || | | |||| |||| ||||
Db    304  CCACCACCAGAGAATC---GAGAAGAGCTATCAATCTGTCAATCCTTTCCGTGTCCGG 248

Qy    349  tcggaataaactttcctgtcttcagcaaatgaagaacccctgctcngtccttgggcgg 408
                ||| | |||||  |||| ||| ||| || ||| | ||| ||| ||| || || ||||
Db    247  GCCGGGTGAGGTTCCCGTGTGGTCAAATTAAGCCGCACTGCCCACTCCTGGGTGGTG 188

Qy    409  gcctt 413
                |||
Db    187  CCCTT 183

RESULT   6
AC011966/c
LOCUS       AC011966      215283 bp    DNA                  HTG        23-DEC-2000
DEFINITION  Homo sapiens chromosome 15 clone RP11-285I14 map 15, WORKING DRAFT
            SEQUENCE    unordered pieces.
ACCESSION   AC011966
VERSION     AC011966  GI:11990705
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 215283)
  AUTHORS   Birren,... Linton,L., Nusbaum,C. and Lander,E.
```

```
VERSION     AC006896.2  GI:4309903
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 298960)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Caehorhabditis elegans clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 298960)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 1, 1999 this sequence version replaced gi:4263452.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1: contig of 3199 bp in length
            *   3200: gap of unknown length
            *   3213: contig of 3199 bp in length
            *   3214: gap of unknown length
            *   5295: contig of 2082 bp in length
            *   5296: gap of unknown length
            *   5308: contig of 2940 bp in length
            *   8249: gap of unknown length
            *   8250: contig of 3261 bp in length
            *   8263: gap of unknown length
            *   8264: contig of 3261 bp in length
            *  11524: gap of unknown length
            *  11525: contig of 3513 bp in length
            *  11538: gap of unknown length
            *  11539: contig of 3513 bp in length
            *  15051: gap of unknown length
            *  15052: contig of 6626 bp in length
            *  15065: gap of unknown length
            *  15066: contig of 6626 bp in length
            *  21691: gap of unknown length
            *  21692: contig of 7600 bp in length
            *  21705: gap of unknown length
            *  21706: contig of 7600 bp in length
            *  29305: gap of unknown length
            *  29306: contig of 10222 bp in length
            *  29319: gap of unknown length
            *  29320: contig of 10222 bp in length
            *  39541: gap of unknown length
            *  39542: contig of 18415 bp in length
            *  39555: gap of unknown length
            *  39956: contig of 18415 bp in length
            *  57970: gap of unknown length
            *  57971: contig of 11911 bp in length
            *  57984: gap of unknown length
            *  57985: contig of 11911 bp in length
            *  69895: gap of unknown length
            *  69896: contig of 20201 bp in length
            *  69909: gap of unknown length
            *  69910: contig of 20201 bp in length
            *  90110: gap of unknown length
            *  90111: contig of 23213 bp in length
            *  90124: gap of unknown length
            *  90125: contig of 23213 bp in length
            * 113337: gap of unknown length
            * 113338: contig of 41255 bp in length
            * 113351: gap of unknown length
            * 113352: contig of 41255 bp in length
            * 154606: gap of unknown length
            * 154607: contig of 110126 bp in length
            * 154620: gap of unknown length
            * 154621: contig of 110126 bp in length
            * 264746: gap of unknown length
            * 264747: contig of 34200 bp in length
            * 264760: gap of unknown length
            * 264761: contig of 34200 bp in length.
FEATURES             Location/Qualifiers
     source          1..298960
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="Y71H2X"
BASE COUNT    96038 a   53465 c   53550 g   95711 t   196 others
ORIGIN

Query Match               8.0%;    Score 37.2;  DB 60;   Length 298960;
Best Local Similarity    53.1%;    Pred. No. 2.9;
Matches   78;  Conservative   0;  Mismatches   69;  Indels   0;  Gaps   0;

Qy    317  ttggccttaaataaacctatctttgtcattcggaataaacttcctgtctcttcagca 376
                || || ||||| || | |||| ||| | || ||| ||| ||| |||| || | ||||
Db 153784  TTTCAAAAAAAATCATTTTTTTGAAATTTTGAAAATATTTCATTTTTTAAACA 153785

Qy    377  aatgaagaacccctgctcngtcctttgggcggcggcctttggacgcgctttgcacgcgctttggacttgtta 436
                ||| | | | || ||| || |||| ||  |  ||  || ||||| || |||| | ||||||
```

TITLE       Homo sapiens chromosome 15, clone RP11-285I14
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 215283)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (17-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 23, 2000 this sequence version replaced gi:7960358.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            --------- Project Information
              Center project name: L2605
              Center clone name: 285_I_14
            --------- Summary Statistics
              Sequencing vector: M13; M77815; 47% of reads
              Sequencing vector: Plasmid; n/a; 53% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 193454 bases at least Q40
              Consensus quality: 200496 bases at least Q30
              Consensus quality: 205965 bases at least Q20
              Insert size: 185000; agarose-fp
              Insert size: 213083; sum-of-contigs
              Quality coverage: 11.3 in Q20 bases; agarose-fp
              Quality coverage: 9.8 in Q20.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1  1012: contig of 1012 bp in length
            *   1013  1112: gap of     100 bp in length
            *   1113  2148: contig of 1036 bp in length
            *   2149  2248: gap of     100 bp in length
            *   2249  3454: contig of 1206 bp in length
            *   3455  3554: gap of     100 bp in length
            *   3555  4747: contig of 1193 bp in length
            *   4748  4847: gap of     100 bp in length
            *   4848  5856: contig of 1009 bp in length
            *   5857  5956: gap of     100 bp in length
            *   5957  7144: contig of 1188 bp in length
            *   7145  7244: gap of     100 bp in length
            *   7245  8271: contig of 1027 bp in length
            *   8272  8371: gap of     100 bp in length
            *   8372  9382: contig of 1011 bp in length
            *   9383  9482: gap of     100 bp in length
            *   9483 10621: contig of 1139 bp in length
            *  10622 10721: gap of     100 bp in length
            *  10722 11851: contig of 1130 bp in length
            *  11852 11951: gap of     100 bp in length
            *  11952 13017: contig of 1066 bp in length
            * 13018 13117: gap of     100 bp in length
            * 13118 14592: contig of 1475 bp in length
            * 14593 14692: gap of     100 bp in length
            * 14693 15932: contig of 1240 bp in length
            * 15933 16032: gap of     100 bp in length
            * 16033 17854: contig of 1822 bp in length
            * 17855 17954: gap of     100 bp in length
            * 17955 19826: contig of 1872 bp in length
            * 19827 19926: gap of     100 bp in length
            * 19927 21064: contig of 1138 bp in length
            * 21065 21164: gap of     100 bp in length
            * 21165 23108: contig of 1944 bp in length
            * 23109 23208: gap of     100 bp in length
            * 23209 24764: contig of 1556 bp in length
            * 24765 24864: gap of     100 bp in length
            * 24865 26747: contig of 1883 bp in length
            * 26748 26847: gap of     100 bp in length
            * 26848 28606: contig of 1759 bp in length
            * 28607 28706: gap of     100 bp in length
            * 28707 30429: contig of 1723 bp in length
            * 30430 30529: gap of     100 bp in length
            * 30530 32780: contig of 2251 bp in length
            * 32781 32880: gap of     100 bp in length
            * 32881 215283: contig of 182403 bp in length.
FEATURES             Location/Qualifiers
     source          1..215283
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-285I14"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1012
                     /note="assembly_fragment"
     misc_feature    1113..2148
                     /note="assembly_fragment"
     misc_feature    2249..3454
                     /note="assembly_fragment"
     misc_feature    3555..4747
                     /note="assembly_fragment"
     misc_feature    4848..5856
                     /note="assembly_fragment"
     misc_feature    5957..7144
                     /note="assembly_fragment"
     misc_feature    7245..8271
                     /note="assembly_fragment"
     misc_feature    8372..9382
                     /note="assembly_fragment"
     misc_feature    9483..10621
                     /note="assembly_fragment"
     misc_feature    10722..11851
                     /note="assembly_fragment"
     misc_feature    11952..13017
                     /note="assembly_fragment"
     misc_feature    13118..14592
                     /note="assembly_fragment"
     misc_feature    14693..15932
                     /note="assembly_fragment"
     misc_feature    16033..17854
                     /note="assembly_fragment"
     misc_feature    17955..19826
                     /note="assembly_fragment"
     misc_feature    19927..21064
                     /note="assembly_fragment"
     misc_feature    21165..23108
                     /note="assembly_fragment"
     misc_feature    23209..24764
                     /note="assembly_fragment"
     misc_feature    24865..26747
                     /note="assembly_fragment"
     misc_feature    26848..28606
                     /note="assembly_fragment"
     misc_feature    28707..30429
                     /note="assembly_fragment"

```
        misc_feature      /note="assembly_fragment"
                          30530..32780
        misc_feature      /note="assembly_fragment".
                          32881..215283
                          /note="assembly_fragment"
BASE COUNT   60660 a   43707 c   46632 g   62078 t   2206 others
ORIGIN

Query Match        7.8%;  Score 36.4;  DB 62;  Length 215283;
Best Local Similarity  61.1%;  Pred. No. 5;
Matches   58;  Conservative   0;  Mismatches   37;  Indels   0;  Gaps   0;

Qy    331 aaacctatcttgtcattcggaataaacttcctgtcttcagcaaatgaagaacccct 390
          | |||||||  ||||||| |||| |||||||||||||| ||||||| | ||| ||
Db  62895 ACACCTATCAACCCTCTTCATTATAAAATCATCTTTCTTTTAGTGACTCAAAATCCACT 62836

C     391 gctcngtccttgggcgggccttttgcacgctttt 425
          || |||||| ||||| || | | | ||| |||
Db  62835 GCCACTTCCTCGGTGAGAGTCTCCACACAACTTT 62801

RESULT   7
AC084469
LOCUS       AC084469    40552 bp   DNA       INV       04-NOV-2000
DEFINITION  Caenorhabditis briggsae cosmid CB045D24, complete sequence.
ACCESSION   AC084469
VERSION     AC084469.1  GI:11094919
KEYWORDS    HTG.
SOURCE      Caenorhabditis briggsae.
ORGANISM    Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 40552)
AUTHORS     Washington University Genome Sequencing Center.
TITLE       The C. briggsae Genome Sequencing Project
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 40552)
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (04-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Genome Sequencing Center
            Submitted by:
                Genome Sequencing Center
                Department of Genetics, Washington University,
                St. Louis, MO 63110, USA
                e-mail: jspieth@watson.wustl.edu

            NOTICE: This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.
FEATURES             Location/Qualifiers
     source          1..40552
                     /organism="Caenorhabditis briggsae"
                     /strain="GujArat G16"
                     /db_xref="taxon:6238"
                     /clone="CB045D24"
BASE COUNT  12516 a   7553 c   7254 g   13229 t
ORIGIN

Query Match        7.8%;  Score 36.2;  DB 4;  Length 40552;
Best Local Similarity  53.6%;  Pred. No. 5.2;
Matches   74;  Conservative   0;  Mismatches   64;  Indels   0;  Gaps   0;

Qy    323 tttaaataaaacctatctttgtcattcggaataaacttcctgtcttcagcaaatgaa 382
          |||||| |||||||| |||||||||||||| ||||  ||  | |||||| ||
Db   3472 TTTAATGAAAACCTACTACTGTATAGGTTACGGTATTAAAAATACCGTATTTCCCAAATAGAA 3531

Qy    383 gaaccctgctcngtccttgggcgggccttttgcacgctttttggacttgttatcaaa 442
          |  | | ||||  ||||| | |||    || | | ||||| | |||||||||
```

```
Db   3532 GCCCACACGAACCGTCATCTTTGAGCTCAACAATCTCGATTTCTGTGAAAGATATCAAAA 3591

Qy    443 aaaccctcctgttcaa 460
          |  ||  ||  || |||
Db   3592 AGAGCCCAACTGATAAAA 3609

RESULT   8
AC013805/c
LOCUS       AC013805   174296 bp   DNA       HTG       16-MAR-2000
DEFINITION  Homo sapiens assembly clone RP11-2003, WORKING DRAFT SEQUENCE, 36 unordered
            pieces.
ACCESSION   AC013805
VERSION     AC013805.4  GI:7249021
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174296)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-2003
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 174296)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,K., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,A.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassilev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141 USA
            On Mar 16, 2000 this sequence version replaced gi:6715925.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------- Project Information
                Center project name: L3956
                Center clone name: 20_O_3
            --------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 142071 bases at least Q40
                Consensus quality: 155814 bases at least Q30
                Consensus quality: 163016 bases at least Q20
                Insert size: 177000; agarose-fp
                Ins: ize: 170796; sum-of-contigs
                Qua.   overage: 2.9 in Q20 bases: agarose-fp
                Qua.   overage: 3.0 in Q20 bases: sum-of-contigs
            -----
            * R:   .... is a 'working draft' sequence. It currently
            * const.  of 36 contigs. The true order of the pieces
            * is not  tten and their order in this sequence record is
            * arbit..  Gaps between the contigs are represented as
            * runs c. t  but the exact sizes of the gaps are unknown.
            * This   .... will be updated with the finished sequence
            * as s it is available and the accession number will
            * be pt.  ved.
```

```
*        1     1366: contig of 1366 bp in length
*     1367     1466: gap of 100 bp in length
*     1467     2696: contig of 1230 bp in length
*     2697     2796: gap of 100 bp in length
*     2797     4009: contig of 1213 bp in length
*     4010     4109: gap of 100 bp in length
*     4110     5925: contig of 1816 bp in length
*     5926     6025: gap of 100 bp in length
*     6026     7794: contig of 1769 bp in length
*     7795     7894: gap of 100 bp in length
*     7895    10197: contig of 2303 bp in length
*    10198    10297: gap of 100 bp in length
*    10298    12776: contig of 2479 bp in length
*    12777    12876: gap of 100 bp in length
*    12877    15103: contig of 2227 bp in length
*    15104    15203: gap of 100 bp in length
*    15204    17869: contig of 2666 bp in length
*    17870    17969: gap of 100 bp in length
*    17970    20373: contig of 2404 bp in length
*    20374    20473: gap of 100 bp in length
*    20474    22164: contig of 1691 bp in length
*    22165    22264: gap of 100 bp in length
*    22265    25326: contig of 3062 bp in length
*    25327    25426: gap of 100 bp in length
*    25427    28477: contig of 3051 bp in length
*    28478    28577: gap of 100 bp in length
*    28578    32211: contig of 3634 bp in length
*    32212    32311: gap of 100 bp in length
*    32312    35480: contig of 3169 bp in length
*    35481    35580: gap of 100 bp in length
*    35581    38760: contig of 3180 bp in length
*    38761    38860: gap of 100 bp in length
*    38861    41874: contig of 3014 bp in length
*    41875    41974: gap of 100 bp in length
*    41975    44720: contig of 2746 bp in length
*    44721    44820: gap of 100 bp in length
*    44821    47765: contig of 2945 bp in length
*    47766    47865: gap of 100 bp in length
*    47866    52024: contig of 4159 bp in length
*    52025    52124: gap of 100 bp in length
*    52125    56524: contig of 4400 bp in length
*    56525    56624: gap of 100 bp in length
*    56625    60455: contig of 3831 bp in length
*    60456    60555: gap of 100 bp in length
*    60556    64084: contig of 3529 bp in length
*    64085    64184: gap of 100 bp in length
*    64185    69030: contig of 4846 bp in length
*    69031    69130: gap of 100 bp in length
*    69131    74003: contig of 4873 bp in length
*    74004    74103: gap of 100 bp in length
*    74104    78646: contig of 4543 bp in length
*    78647    78746: gap of 100 bp in length
*    78747    85684: contig of 6938 bp in length
*    85685    85784: gap of 100 bp in length
*    85785    92158: contig of 6374 bp in length
*    92159    92258: gap of 100 bp in length
*    92259    98076: contig of 5818 bp in length
*    98077    98176: gap of 100 bp in length
*    98177   105628: contig of 7452 bp in length
*   105629   105728: gap of 100 bp in length
*   105729   115493: contig of 9765 bp in length
*   115494   115593: gap of 100 bp in length
*   115594   124340: contig of 8747 bp in length
*   124341   124440: gap of 100 bp in length
*   124441   132343: contig of 7903 bp in length
*   132344   132443: gap of 100 bp in length
*   132444   144952: contig of 12509 bp in length
*   144953   145052: gap of 100 bp in length
*   145053   158986: contig of 13934 bp in length
*   158987   159086: gap of 100 bp in length
*   159087   174296: contig of 15210 bp in length.
FEATURES             Location/Qualifiers
     source          1 ..174296
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1366
                     /note="assembly_fragment"
     misc_feature    1467..2696
                     /note="assembly_fragment"
     misc_feature    2797..4009
                     /note="assembly_fragment"
     misc_feature    4110..5925
                     /note="assembly_fragment"
     misc_feature    6026..7794
                     /note="assembly_fragment"
     misc_feature    7895..10197
                     /note="assembly_fragment"
     misc_feature    10298..12776
                     /note="assembly_fragment"
     misc_feature    12877..15103
                     /note="assembly_fragment"
     misc_feature    15204..17869
                     /note="assembly_fragment"
     misc_feature    17970..20373
                     /note="assembly_fragment"
     misc_feature    20474..22164
                     /note="assembly_fragment"
     misc_feature    22265..25326
                     /note="assembly_fragment"
     misc_feature    25427..28477
                     /note="assembly_fragment"
     misc_feature    28578..32211
                     /note="assembly_fragment"
     misc_feature    32312..35480
                     /note="assembly_fragment"
     misc_feature    35581..38760
                     /note="assembly_fragment"
     misc_feature    38861..41874
                     /note="assembly_fragment"
     misc_feature    41975..44720
                     /note="assembly_fragment"
     misc_feature    44821..47765
                     /note="assembly_fragment"
     misc_feature    47866..52024
                     /note="assembly_fragment"
     misc_feature    52125..56524
                     /note="assembly_fragment"
     misc_feature    56625..60455
                     /note="assembly_fragment"
     misc_feature    60556..64084
                     /note="assembly_fragment"
     misc_feature    64185..69030
                     /note="assembly_fragment"
     misc_feature    69131..74003
                     /note="assembly_fragment"
     misc_feature    74104..78646
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    78747..85684
                     /note="assembly_fragment"
     misc_feature    85785..92158
                     /note="assembly_fragment"
     misc_feature    92259..98076
```

```
Query Match    7.7%;  Score 36;  DB 63;  Length 174296;
Best Local Similarity   50.9%;  Pred. No. 6.6;
Matches  84;  Conservative   0;  Mismatches  81;  Indels   0;  Gaps   0;

Qy     287 tatcaccctacccacaccccgtaccaaccttgcccttaaataaacctatcttgtca 346
           ||||| | |||||||| |||||| ||||||| | ||||| || ||| || |||
Db   13839 TATCTCATTCCCCACACCGTTTCACAAACTTTAGTTTCTGATAAAGGGAATCAGGCTCA 13780

Qy     347 tttcggaataaacttcctgtctttcagcaaatgaagaacccctgctcngtcctttgggc 406
```

```
Db   13779 ATGTCACTTANGTTTCACTATCTCTCGGGCCTTTGCAAATGCTGTTCCCTCCATTTGGA 13720
              |  |  ||   ||  || | ||| | |||   | ||   ||| |||  || ||  |  ||
Qy     407 gggccttttgcacgcttttgacttgttatcaaaaaacccttc 451
              |||  |||            ||  ||| |    ||  | |||
Db   13719 CAGCTTTTCCTTCAGCTGGCACTTGGAAATCCTAAATCCATTC 13675
```

```
RESULT     9
AC026673
LOCUS       AC026673    334796 bp    DNA           HTG       15-NOV-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-56B20, WORKING DRAFT SEQUENCE,
            70 unordered pieces.
ACCESSION   AC026673
VERSION     AC026673.12  GI:10047500
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 334796)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks.S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie.S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,D., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 334796)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 10, 2000 this sequence version replaced gi:9690210.
            -------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
```

```
               Contact: hgsc-help@bcm.tmc.edu
            -------- Project Information
               Center project name: HAOX
               Center clone name: RP11-56B20
            -------- Summary Statistics
               Sequencing vector: M13; L08821
               Chemistry: Dye-primer Bodipy: 43% of reads
               Chemistry: Dye-terminator Big Dye: 57% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 282951 bases at least Q40
               Consensus quality: 301471 bases at least Q30
               Consensus quality: 312551 bases at least Q20
               Estimated insert size: 306588; sum-of-contigs estimation
               Quality coverage: 0x in Q20 bases; agarose-fp estimation
               Quality coverage: 3.1x in Q20 bases; sum-of-contigs estimation

          * NOTE: Estimated insert size may differ from sequence length
          *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
          * NOTE: This sequence may represent more than one clone.
          * NOTE: This is a 'working draft' sequence. It currently
          *       consists of 70 contigs. The true order of the pieces
          *       is not known and their order in this sequence record is
          *       arbitrary. Gaps between the contigs are represented as
          *       runs of N, but the exact sizes of the gaps are unknown.
          *       This record will be updated with the finished sequence
          *       as soon as it is available and the accession number will
          *       be preserved.
          *        1:        contig of 50044 bp in length
          *    50045:        gap of unknown length
          *    50144:        contig of 30382 bp in length
          *    80526:        gap of unknown length
          *    80527:        contig of 26117 bp in length
          *   106743:        gap of unknown length
          *   106744:        contig of 24935 bp in length
          *   131778:        gap of unknown length
          *   131779:        contig of 14087 bp in length
          *   145865:        gap of unknown length
          *   145966:        contig of 11458 bp in length
          *   146066:        gap of unknown length
          *   157523:        contig of 10723 bp in length
          *   157624:        gap of unknown length
          *   168346:        contig of 5181 bp in length
          *   168446:        gap of unknown length
          *   168447:        contig of 5733 bp in length
          *   173627:        gap of unknown length
          *   173628:        contig of 7480 bp in length
          *   173727:        gap of unknown length
          *   173728:        contig of 5927 bp in length
          *   179460:        gap of unknown length
          *   179461:        contig of 5651 bp in length
          *   179560:        gap of unknown length
          *   179561:        contig of 5605 bp in length
          *   187040:        gap of unknown length
          *   187041:        contig of 4805 bp in length
          *   187140:        gap of unknown length
          *   187141:        contig of 2864 bp in length
          *   193067:        gap of unknown length
          *   193068:        contig of 3472 bp in length
          *   193167:        gap of unknown length
          *   193168:        contig of 4613 bp in length
          *   198818:        gap of unknown length
          *   198819:        contig of 4213 bp in length
          *   198918:        gap of unknown length
          *   204523:        contig of 3020 bp in length
          *   204524:        gap of unknown length
          *   204623:        contig of 3347 bp in length
          *   204624:        gap of unknown length
          *   209428:        contig of 4432 bp in length
          *   209429:        gap of unknown length
          *   209528:        contig of 3472 bp in length
          *   209529:        gap of unknown length
          *   212392:        contig of 3472 bp in length
          *   212393:        gap of unknown length
          *   212492:        contig of 2523 bp in length
          *   212493:        gap of unknown length
          *   217105:        gap of unknown length
          *   217106:        contig of 2523 bp in length
          *   217205:
          *   217206:
          *   222756:
          *   222757:
          *   222856:
          *   222857:
          *   226203:
          *   226303:
          *   226  :
          *   230735:
          *   230835:
          *   234307:
          *   234407:
          *   234  :
          *   234  :
          *   238620:
          *   238720:
          *   241740:
          *   241840:
          *   244363:
          *   244463:
```

```
Qy    347  tttcggaataaacttcctgtcttcagcaaatgaagaaccccgtcngtccttggc  406
           ||  ||| |||| |||| ||||| ||| ||| || |||| |||||| ||| ||||
Db 109725  ATGTCACTTATGTTCACTATCTCCTGGGCCTTTGCAAATGCTGTTCCCTCCATTTGGA  109784

Qy    407  gggccttttgcacgctttggacttgttatcaaaaaaacccttc  451
           |||| ||    |||| || ||||| ||| |||| |||||||||
Db 109785  CAGCTTTCCTTTCAGCGTGCACTTGGAAATCCTAAATCCATTC  109829
```

```
RESULT 10
AL365272
LOCUS       AL365272        132449 bp    DNA             PRI       23-DEC-2000
DEFINITION  Human DNA sequence from clone RP11-328C17 on chromosome 6, complete
            sequence.
ACCESSION   AL365272
VERSION     AL365272.20  GI:11991402
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 132449)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 24, 2000 this sequence version replaced gi:11875910.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the overlapping clone
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements. Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-328C17 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-328C17 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP11-328C17 is at 132449 in this
            sequence. The true left end of clone RP11-233K4 is at 64893 in this
            sequence. The true right end of clone RP3-416J7 is at 100 in this
            sequence. The true right end of clone RP1-125A24 is at 61243 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1..132449
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-328C17"
                     /clone_lib="RPCI-11.2"
     repeat_region   1..162
                     /note="L1M4 repeat: matches 4133..5719 of consensus"
     repeat_region   1563..1641
                     /note="MLT1A2 repeat: matches 1..80 of consensus"
     repeat_region   1669..2199
                     /note="L1M4 repeat: matches 3396..3937 of consensus"
```

```
 *     244464   247852:  contig  of 3389  bp in length
 *     247853   247952:  gap of  unknown  length
 *     247953   251204:  contig  of 3252  bp in length
 *     251205   251304:  gap of  unknown  length
 *     251305   253924:  contig  of 2620  bp in length
 *     253925   254024:  gap of  unknown  length
 *     254025   258689:  contig  of 4665  bp in length
 *     258690   258789:  gap of  unknown  length
 *     258790   261355:  contig  of 2566  bp in length
 *     261356   261455:  gap of  unknown  length
 *     261456   263987:  contig  of 2532  bp in length
 *     263988   264087:  gap of  unknown  length
 *     264088   267491:  contig  of 3404  bp in length
 *     267492   267591:  gap of  unknown  length
 *     267592   269625:  contig  of 2034  bp in length
 *     269626   269725:  gap of  unknown  length
 *     269726   271829:  contig  of 2104  bp in length
 *     271830   271929:  gap of  unknown  length
 *     271930   274961:  contig  of 3032  bp in length
 *     274962   275061:  gap of  unknown  length
 *     275062   277232:  contig  of 2171  bp in length
 *     277233   277332:  gap of  unknown  length
 *     277333   278979:  contig  of 1646  bp in length
 *     278979   279078:  gap of  unknown  length
 *     279079   281944:  contig  of 2866  bp in length
 *     281945   282044:  gap of  unknown  length
 *     282045   284703:  contig  of 2659  bp in length
 *     284704   284803:  gap of  unknown  length
 *     284804   286499:  contig  of 1696  bp in length
 *     286500   286599:  gap of  unknown  length
 *     286600   289303:  contig  of 2704  bp in length
 *     289304   289403:  gap of  unknown  length
 *     289404   291372:  contig  of 1969  bp in length
 *     291373   291472:  gap of  unknown  length
 *     291472   293293:  contig  of 1821  bp in length
 *     293294   293393:  gap of  unknown  length
 *     293394   295064:  contig  of 1671  bp in length
 *     295065   295164:  gap of  unknown  length
 *     295164   296870:  contig  of 1706  bp in length
 *     296871   296970:  gap of  unknown  length
 *     296970   298568:  contig  of 1598  bp in length
 *     298568   298669:  gap of  unknown  length
 *     298669   299734:  contig  of 1066  bp in length
 *     299734   299834:  gap of  unknown  length
 *     299835   301632:  contig  of 1798  bp in length
 *     301633   301732:  gap of  unknown  length
 *     301732   303247:  contig  of 1515  bp in length
 *     303247   303348:  gap of  unknown  length
 *     303348   304660:  contig  of 1313  bp in length
 *     304660   304760:  gap of  unknown  length
 *     304761   306454:  contig  of 1694  bp in length
 *     306455   306554:  gap of  unknown  length
 *     306555   307631:  contig  of 1076  bp in length
 *     307631   307730:  gap of  unknown  length
 *     307731   309334:  contig  of 1604  bp in length
 *     309335   309434:  gap of  unknown  length
 *     309435   310606:  contig  of 1172  bp in length
 *     310607   310706:  gap of  unknown  length
 *     310706   311844:  contig  of 1138  bp in length
 *     311845   311944:  gap of  unknown  length
 *     311945   313346:  contig  of 1402  bp in length
 *     313347   313446:  gap of  unknown  length
 *     313446   315156:  contig  of 1710  bp in length
 *     315157   315256:  gap of  unknown  length
```

```
Query Match             7.7%;    Score 36;  DB 70;  Length 334796;
Best Local Similarity  50.9%;    Pred. NO. 6.9;
Matches   84;  Conservative   0;  Mismatches  81;  Indels  0;  Gaps  0;

Qy    287  tatcaccctacccaccaccccgtgaccaacctggccttaaataaaaacctatctttgtca  346
           |||||||||||   ||||||||  ||| ||||  |||| |||| |  |||| ||||| |||
Db 109665  TATCTCATTCCCCACACACCGTTCACAAACTTTAGTTTCTGATAAAGGGGAATCAGGCTCA  109724
```

```
repeat_region   2200..2474
repeat_region   2475..4279
                /note="AluSx repeat: matches 38. .312 of consensus"
repeat_region   4306..4589
                /note="L1M4 repeat: matches 1686. .3396 of consensus"
misc_feature    /note="L1M4 repeat: matches 1254. .1544 of consensus"
                complement(4704..5324)
                /note="match: GSS: Em:AQ628183"
repeat_region   4799..4938
                /note="L1M4c repeat: matches 785. .924 of consensus"
misc_feature    complement(4809..5326)
                /note="match: GSS: Em:AQ774852"
misc_feature    complement(4860..5272)
                /note="match: GSS: Em:AQ787689"
repeat_region   4943..5094
misc_feature    /note="L1M4c repeat: matches 272. .420 of consensus"
                complement(5020..5329)
                /note="match: GSS: Em:AQ494601"
repeat_region   5142..5435
                /note="MLT1D repeat: matches 170. .428 of consensus"
repeat_region   5436..5744
                /note="AluSx repeat: matches 1. .310 of consensus"
repeat_region   5745..5869
                /note="MLT1D repeat: matches 428. .568 of consensus"
repeat_region   6450..6686
                /note="MLT1D repeat: matches 7. .261 of consensus"
repeat_region   6916..7994
                /note="L1M4 repeat: matches 3582. .4688 of consensus"
repeat_region   7984..8050
                /note="L1MB8 repeat: matches 5808. .5884 of consensus"
repeat_region   8064..8169
                /note="L1M4 repeat: matches 3479. .3593 of consensus"
repeat_region   8178..8224
                /note="L1MB4 repeat: matches 6124. .6169 of consensus"
repeat_region   8225..8525
                /note="AluSx repeat: matches 1. .301 of consensus"
repeat_region   8526..8915
                /note="L1MB4 repeat: matches 5725. .6124 of consensus"
misc_feature    complement(8862..9353)
                /note="match: GSS: Em:AQ355497"
repeat_region   9029..9150
                /note="2 copies 61 mer 95% conserved"
misc_feature    complement(9052..9484)
                /note="match: GSS: Em:AQ175845"
misc_feature    9388..9675
                /note="match: GSS: Em:AQ135206"
misc_feature    9487..9840
                /note="match: GSS: Em:AQ717684"
misc_feature    9491..9932
                /note="match: GSS: Em:AQ147759"
repeat_region   9524..9697
                /note="2 copies 49 mer 91% conserved"
repeat_region   10046..10143
                /note="MER53 repeat: matches 1. .189 of consensus"
repeat_region   11168..11361
                /note="MLT1B repeat: matches 101. .309 of consensus"
repeat_region   11448..11756
                /note="AluY repeat: matches 1. .309 of consensus"
repeat_region   12109..12515
                /note="L2 repeat: matches 1747. .2155 of consensus"
repeat_region   12798..12947
                /note="L2 repeat: matches 29. .206 of consensus"
repeat_region   12975..13026
                /note="MIR repeat: matches 186. .237 of consensus"
repeat_region   13005..13103
                /note="L2 repeat: matches 2417. .2516 of consensus"
repeat_region   13331..13664
                /note="MLT1B repeat: matches 4. .361 of consensus"
repeat_region   13704..13907
                /note="LTR33 repeat: matches 297. .511 of consensus"
repeat_region   15450..15530
                /note="L2 repeat: matches 2668. .2749 of consensus"
repeat_region   15876..15935
```

```
repeat_region   16202..16294
                /note="2 copies 30 mer 100% conserved"
misc_feature    17335..17668
                /note="MER81 repeat: matches 2. .94 of consensus"
repeat_region   17419..17715
                /note="match: GSS: Em:AQ104665"
misc_feature    17995..18362
                /note="AluSx repeat: matches 1. .296 of consensus"
repeat_region   19374..19496
                /note="match: GSS: Em:AQ151578"
repeat_region   20271..20430
                /note="L2 repeat: matches 2579. .2709 of consensus"
repeat_region   21095..21389
                /note="MER20 repeat: matches 55. .218 of consensus"
repeat_region   21830..21946
                /note="AluSx repeat: matches 1. .295 of consensus"
repeat_region   22328..22772
                /note="MIR repeat: matches 42. .152 of consensus"
repeat_region   22768..22848
                /note="L2 repeat: matches 2242. .2710 of consensus"
misc_feature    23314..24204
                /note="L2 repeat: matches 2661. .2750 of consensus"
misc_feature    23596..23601
                /note="CpG island"
                /evidence=not_experimental
repeat_region   25664..25783
                /note="Weak data."
repeat_region   25786..26131
                /note="L2 repeat: matches 2588. .2741 of consensus"
repeat_region   26858..27183
                /note="MER33 repeat: matches 1. .324 of consensus"
repeat_region   27208..27334
                /note="AluJb repeat: matches 1. .301 of consensus"
misc_feature    27673..27833
                /note="MIR repeat: matches 81. .221 of consensus"
misc_feature    complement(28112..28516)
                /note="MIR repeat: matches 7. .175 of consensus"
repeat_region   28535..29088
                /note="match: GSS: Em:AQ212338"
misc_feature    28560..28684
                /note="match: GSS: Em:B51862"
misc_feature    28588..28828
                /note="MIR repeat: matches 79. .210 of consensus"
repeat_region   28598..28841
                /note="match: GSS: Em:B78941"
misc_feature    28849..29048
                /note="match: GSS: Em:B88272"
repeat_region   30148..30232
                /note="MER20 repeat: matches 2. .214 of consensus"
misc_feature    complement(31265..31862)
                /note="L2 repeat: matches 2610. .2692 of consensus"
misc_feature    complement(31333..31862)
                /note="match: GSS: Em:AQ625926"
misc_feature    complement(31491..31857)
                /note="match: GSS: Em:AQ625934"
repeat_region   31733..31875
                /note="L2 repeat: matches 2039. .2197 of consensus"
repeat_region   33899..33992
                /note="match: GSS: Em:AQ131221"
repeat_region   34111..34427
                /note="L2 repeat: matches 2578. .2667 of consensus"
repeat_region   35300..35435
                /note="MER31A repeat: matches 4. .477 of consensus"
                /note="MER5A repeat: matches 121. .189 of consensus"
```

```
Query Match          7.7%;    Score 35.8;   DB 90;   Length 132449;
Best Local Similarity  52.3%;  Pred No. 7.4;
Matches  79;   Mismatches  0;   Indels  0;   Gaps  0;

Qy    214 ggccctga tttctcttaccattgtcggattaccgctatggcgcttcatcaaca 273
          |||  ||  | ||||| ||| || |  | | ||| ||  |   || | |||
Db 115574 GGCTCTG ..AGCACTTAGGAAAGAAGGATGTATGCCCTGGGCTGCTGGGAGCACT 115633
```

```
Qy    274  gaggatgccctagtatcacccctaccacacccgtgaccaacccttggccttaaataaa  333
           |-|| |-| ||| ||||| | |-| |||       |-||| || ||||-||   |||||
Db 115634  GTGGGAACTCTGGGATCCCCTCCTCCTAAGGGCTGTGCTACCCTGATTTGGAAACAATA  115693

Qy    334  ctatctttgtcattcggaataaactttcc  364
           |-|-||||   |-|| ||| |||||-|||
Db 115694  CAGAGGCTTCTCTCCTGCAAGAAAACATGCC  115724
```

```
RESULT 11
AC058816
LOCUS       AC058816    171350 bp    DNA           HTG       17-AUG-2000
DEFINITION  Homo sapiens chromosome 6 clone RP11-233K4, WORKING DRAFT SEQUENCE,
            21 unordered pieces.
ACCESSION   AC058816
VERSION     AC058816.3  GI:9838113
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171350)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171350)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 17, 2000 this sequence version replaced gi:7709945.

            -------------- Genome Center --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            ----------- Project Information -----------
            Center project name: H_NH0233K04
            ----------- Summary Statistics -----------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 161421 bases at least Q40
            Consensus quality: 164672 bases at least Q30
            Consensus quality: 166316 bases at least Q20
            Insert size: 187000; agarose-fp
            Insert size: 170998; sum-of-contigs
            Quality coverage: 4.43 in Q20 bases; agarose-fp
            Quality coverage: 4.92 in Q20 bases; sum-of-contigs

            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1   2122: contig of 2122 bp in length
            *   2123   2222: gap of unknown length
            *   2223   3812: contig of 1590 bp in length
            *   3813   3912: gap of unknown length
            *   3913   5636: contig of 1724 bp in length
            *   5637   5736: gap of unknown length
            *   5737   8302: contig of 2566 bp in length
            *   8303   8402: gap of unknown length
            *   8403  13037: contig of 4635 bp in length
            *  13038  13137: gap of unknown length
            *  13138  16443: contig of 3306 bp in length
            *  16444  16543: gap of unknown length
            *  16544  19852: contig of 3309 bp in length
            *  19853  19952: gap of unknown length
            *  19953  22830: contig of 2878 bp in length
            *  22831  22930: gap of unknown length
            *  22931  27991: contig of 5061 bp in length
            *  27992  28091: gap of unknown length
            *  28092  34326: contig of 6235 bp in length
            *  34327  34426: gap of unknown length
            *  34427  39247: contig of 4821 bp in length
            *  39248  39347: gap of unknown length
            *  39348  48704: contig of 9357 bp in length
            *  48705  48804: gap of unknown length
            *  48805  56906: contig of 8102 bp in length
            *  56907  57006: gap of unknown length
            *  57007  66014: contig of 9008 bp in length
            *  66015  66114: gap of unknown length
            *  66115  75467: contig of 9353 bp in length
            *  75468  75567: gap of unknown length
            *  75568  86090: contig of 10523 bp in length
            *  86091  86190: gap of unknown length
            *  86191 103375: contig of 17185 bp in length
            * 103376 103475: gap of unknown length
            * 103476 119595: contig of 16120 bp in length
            * 119596 119695: gap of unknown length
            * 119696 135009: contig of 15314 bp in length
            * 135010 135109: gap of unknown length
            * 135110 152926: contig of 17817 bp in length
            * 152927 153026: gap of unknown length
            * 153027 171350: contig of 18324 bp in length.
FEATURES             Location/Qualifiers
     source          1..171350
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-233K4"
     misc_feature    1..2122
                     /note="assembly_name:Contig10"
     misc_feature    2223..3812
                     /note="assembly_name:Contig11"
     misc_feature    3913..5636
                     /note="assembly_name:Contig12"
     misc_feature    5737..8302
                     /note="assembly_name:Contig13"
     misc_feature    8403..13037
                     /note="assembly_name:Contig14"
     misc_feature    13138..16443
                     /note="assembly_name:Contig15"
     misc_feature    16544..19852
                     /note="assembly_name:Contig16"
     misc_feature    19953..22830
                     /note="assembly_name:Contig17
                     clone_end:SP6
                     vector_side:right"
     misc_feature    22931..27991
                     /note="assembly_name:Contig18"
     misc_feature    28092..34326
                     /note="assembly_name:Contig19"
     misc_feature    34427..39247
                     /note="assembly_name:Contig20"
     misc_feature    39348..48704
                     /note="assembly_name:Contig21"
     misc_feature    48805..56906
                     /note="assembly_name:Contig22"
     misc_feature    57007..66014
                     /note="assembly_name:Contig23"
     misc_feature    66115..75467
                     /note="assembly_name:Contig24"
     misc_feature    75568..86090
                     /note="assembly_name:Contig25"
     misc_feature    86191..103375
                     /note="assembly_name:Contig26
                     clone_end:T7"
```

```
     misc_feature       vector_side:left"
                        103476..119595
     misc_feature       119696..135009
                        /note="assembly_name:Contig27"
     misc_feature       135110..152926
                        /note="assembly_name:Contig28"
     misc_feature       153027..171350
                        /note="assembly_name:Contig29"
                        /note="assembly_name:Contig30"
BASE COUNT    44302 a    39607 c   39553 g   45877 t     2011 others
ORIGIN

Query Match                  7.7%;     Score 35.8;   DB 72;   Length 171350;
Best Local Similarity   52.3%;     Pred. No. 7.6;
Matches   79;  Conservative    0; Mismatches   72; Indels    0;   Gaps    0;

C        214 ggccctgatcggatttcttaccattgtcggattaccgctatggcgcttcatcaaca 273
            ||| ||| ||| ||| | |||| | | |||| | | ||||| | ||| | ||| ||
Db 164544 GGCTCTGCCCGGTAGCACTTAGGAAAGAAGGATGTATGCCCTGGCGTGCTGGGAGCACT 164603

Qy        274 gaggatgcccctagtatcaccctacccacaccccgtgaccaaccttggccttaaataaa 333
            |  ||| | ||| ||| | ||| | | |||  ||   | | || ||| |||| |   |
Db 164604 GTGGGAACTCTGGGATCCCCCTCCTCCTCCTTAAGGGCTGTGCTACCCTGATTTGGAAACAATA 164663

Qy        334 cctatctttgtcattcggaataaacttcc 364
            | ||| | | ||||| |    |||| ||
Db 164664 CAGAGGCTTCTCTCCTGCAAGAAAACATGCC 164694


RESULT 12
AC044842
LOCUS       AC044842  172307 bp   DNA                HTG      25-JUN-2000
DEFINITION  Homo sapiens chromosome 2 clone RP11-178E20 map 2, WORKING DRAFT
            SEQUENCE, 23 unordered pieces.
ACCESSION   AC044842
VERSION     AC044842.2  GI:8705156
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172307)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 2, clone RP11-178E20
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 172307)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Coiangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Lockk,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wynan,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 25, 2000 this sequence version replaced gi:7543808.
            All repeats were identified using RepeatMasker:
```

```
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
------------------- Genome Center
        Center: Whitehead Institute/ MIT Center for Genome Research
        Center code: WIBR
        Web site: http://www-seq.wi.mit.edu
        Contact: sequence_submissions@genome.wi.mit.edu
------------------- Project Information
        Center project name: L9002
        Center clone name: 178_E_20
------------------- Summary Statistics
        Sequencing vector: M13; M77815; 100% of reads
        Chemistry: Dye-terminator Big Dye; version 0.960731
        Assembly program: Phrap; version 0.960731
        Consensus quality: 156689 bases at least Q40
        Consensus quality: 164477 bases at least Q30
        Consensus quality: 167863 bases at least Q20
        Insert size: 171000: agarose-fp
        Insert size: 170107: sum-of-contigs
        Quality coverage: 3.7 in Q20 bases: agarose-fp
        Quality coverage: 3.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*               1        19: contig of 19 bp in length
*              20      119: gap of     100 bp
*             120     1921: contig of 1802 bp in length
*            1922     2021: gap of     100 bp
*            2022     3689: contig of 1668 bp in length
*            3690     3789: gap of     100 bp
*            3790     6531: contig of 2742 bp in length
*            6532     6631: gap of     100 bp
*            6632     8134: contig of 1503 bp in length
*            8135     8234: gap of     100 bp
*            8235    11919: contig of 3685 bp in length
*           11920    12019: gap of     100 bp
*           12020    15244: contig of 3225 bp in length
*           15245    15344: gap of     100 bp
*           15345    17344: contig of 2000 bp in length
*           17345    17444: gap of     100 bp
*           17445    21830: contig of 4386 bp in length
*           21831    21930: gap of     100 bp
*           21931    25081: contig of 3151 bp in length
*           25082    25181: gap of     100 bp
*           25182    29736: contig of 4555 bp in length
*           29737    29836: gap of     100 bp
*           29837    36853: contig of 7017 bp in length
*           36854    36953: gap of     100 bp
*           36954    41270: contig of 4317 bp in length
*           41271    41370: gap of     100 bp
*           41371    46765: contig of 5395 bp in length
*           46766    46865: gap of     100 bp
*           46866    57601: contig of 10736 bp in length
*           57602    57701: gap of     100 bp
*           57702    67879: contig of 10178 bp in length
*           67880    67979: gap of     100 bp
*           6     78235: contig of 10256 bp in length
*           78335: gap of     100 bp
*           89176: contig of 10841 bp in length
*           49276: gap of     100 bp
*          106081: contig of 16805 bp in length
*          10    21387: contig of 15206 bp in length
*          12  121487: gap of     100 bp
*          12  137207: contig of 15720 bp in length
*          137   137307: gap of     100 bp
*          137  53745: contig of 16438 bp in length
```

```
Db 164468  GTGGGAACTTCGGGATCCCCTCCTCCTCCTAAGGCGTGTGTGCTACCCTGATTTGGAAACAATA 164527
           |||   |  | | | || || ||||| ||  |  | | | |   ||| || || |
Qy    334  cctatctttgtcattcggaataacttcc 364
           | ||       | ||  || |  |   ||||
Db 164528  CAGAGGCTTCTCTCCTGCAAGAAACATGCC 164558
```

RESULT 13
AL589962/c

| LOCUS | AL589962 | 182529 bp | DNA | HTG | 17-MAR-2001 |

DEFINITION  Homo sapiens chromosome 6 clone RP11-233K4, *** SEQUENCING IN
            PROGRESS ***, 6 unordered pieces.
ACCESSION   AL589962
VERSION     AL589962.1  GI:13396817
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182529)
AUTHORS     Almeida,J.
TITLE       Direct Submission
JOURNAL     Submitted (15-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA233K4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 28% of reads
            Sequencing vector: plasmid; L08752; 71% of reads
            Chemistry: Dye-terminator Big Dye; 73% of reads
            Chemistry: Dye-primer-amersham; 26% of reads
            Chemistry: Dye-primer Big Dye; 0% of reads
            Consensus quality: 181426 bases at least Q40
            Consensus quality: 181607 bases at least Q30
            Consensus quality: 181768 bases at least Q20
            Insert size: 182029; sum-of-contigs
            Insert size: 168064; 10.5% error; agarose-fp
            Quality coverage: 14.10x in Q20 bases; sum-of-contigs Quality
            coverage: 15.93x in Q20 bases; agarose-fp
            --------------
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1  81014: contig of 81014 bp in length
            *    81015  81114: gap of      100 bp
            *    81115 105285: contig of 24171 bp in length
            *   105286 105385: gap of      100 bp
            *   105386 119420: contig of 14035 bp in length
            *   119421 119520: gap of      100 bp
            *   119521 141236: contig of 21716 bp in length
            *   141237 141336: gap of      100 bp
            *   141337 180096: contig of 38760 bp in length
            *   180097 180196: gap of      100 bp
            *   180197 182529: contig of 2333 bp in length
FEATURES             Location/Qualifiers
     source          1..182529
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-233K4"
```

---

```
     *     153746 153845: gap of     100 bp
     *     153846 172307: contig of 18462 bp in length.
FEATURES             Location/Qualifiers
     source          1..172307
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-178E20"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..19
                     /note="assembly_fragment"
                     clone_end:T7
                     vector_side:left"
     misc_feature    120..1921
                     /note="assembly_fragment"
     misc_feature    2022..3689
                     /note="assembly_fragment"
     misc_feature    3790..6531
                     /note="assembly_fragment"
     misc_feature    6632..8134
                     /note="assembly_fragment"
     misc_feature    8235..11919
                     /note="assembly_fragment"
     misc_feature    12020..15244
                     /note="assembly_fragment"
     misc_feature    15345..17344
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    17445..21830
                     /note="assembly_fragment"
     misc_feature    21931..25081
                     /note="assembly_fragment"
     misc_feature    25182..29736
                     /note="assembly_fragment"
     misc_feature    29837..36853
                     /note="assembly_fragment"
     misc_feature    36954..41270
                     /note="assembly_fragment"
     misc_feature    41371..46765
                     /note="assembly_fragment"
     misc_feature    46866..57601
                     /note="assembly_fragment"
     misc_feature    57702..67879
                     /note="assembly_fragment"
     misc_feature    67980..78235
                     /note="assembly_fragment"
     misc_feature    78336..89176
                     /note="assembly_fragment"
     misc_feature    89277..106081
                     /note="assembly_fragment"
     misc_feature    106182..121387
                     /note="assembly_fragment"
     misc_feature    121488..137207
                     /note="assembly_fragment"
     misc_feature    137308..153745
                     /note="assembly_fragment"
     misc_feature    153846..172307
                     /note="assembly_fragment"
BASE COUNT  44249 a  40378 c  41592 g  43880 t    2208 others
ORIGIN
```

Query Match         7.7%;     Score 35.8;  DB 71;  Length 172307;
Best Local Similarity  52.3%;  Pred. No. 7.6;
Matches   79;  Conservative   0;  Mismatches   72;  Indels   0;  Gaps   0;

```
Qy    214  ggccctgatcggattctcttaccattgtcggattaccgctatggcgcttcatcaaca 273
           ||| ||| ||| |  | ||| | ||||| |  ||||| || |  || |  |   ||||
Db 164408  GGCTCTGCCCGGTAGCACTTAGGAAAGAAGGATGTATGCCCTGGGCTGCTGGGAGCACT 164467
                                                   ||
Qy    274  gaggatgcccagtatcaccctacccacccccgtgaccaacctttgccttaaataaa 333
```

```
                    Center code: UWMSC
                    Web site: http://chroma.mbt.washington.edu/msg_www
                    Contact: leerowen@systemsbiology.org
                    ---------- Summary Statistics
                    Sequencing vector: pUC18; L08752
                    Chemistry: Dye-terminator Big Dye; 90% of reads
                    Chemistry: Dye-primer Big Dye; 10% of reads
                    Assembly program: Phrap; version 0.990399
                    Note:data from AC068714 [Drafting center UWMSC], and AC068722
                    [Drafting center UWMSC] were added for finishing.
FEATURES            Location/Qualifiers
     source         1..180538
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="15"
                    /map="15q21.1"
                    /clone="RP11-718O11"
                    /clone_lib="RPCI human BAC library 11"
                    /note="This clone overlaps CTD-2306A12 AC068722,and
                    RP11-31508 AC068714 . Data from overlapping BACs were
                    combined and the consensus sequence determined from
                    RP11-718O11 to the extent possible."
     misc_feature   1..42908
                    /note="overlap with CTD-2306A12 AC068722"
     unsure         99323..99339
                    /note="low quality data"
     unsure         123136..123178
                    /note="low quality data"
     unsure         135789..135793
                    /note="low quality data"
     unsure         138584..138592
                    /note="low quality data"
     unsure         149401..149447
                    /note="low quality data"
     unsure         162342..162350
                    /note="low quality data"
     unsure         166805..166809
                    /note="low quality data"
     misc_feature   174144..180538
                    /note="overlap with RP11-31508 AC068714"
BASE COUNT     51952 a   35579 c  37307 g  55700 t
ORIGIN


Query Match                     7.6%;     Score 35.2;   DB 88;   Length 180538;
Best Local Similarity   54.3%;     Pred. No. 12;
Matches    70;   Conservative     0;   Mismatches     59;   Indels     0;   Gaps     0;

Qy     329 taaacctatctttgtcatttcggaataaacttcctgtctttcagcaaatgaagaaccc 388
           |||||||||  ||| |||||||| ||| |||||||| |||| || ||||| |||||| |
Db   79773 TGAATTCTTTCTCTGGCAATCCAGAGATTTTGTCTGGGTTGTATCCATTGCTCGTTAG 79714

Qy     389 ctgctcngtcctttggcgggcggcctttcgcacgctttatcaaaaaccc 448
           || ||||| ||| ||| ||| |||| |||||||| ||||||||||| |||
Db   79713 CTAGTGTGATCTTTTGGGAGTGCTAAAGAACTTTGTTTTGCCATGTATCAGAATTATTT 79654

Qy     449 ttcttgttc 457
           |||||||||
Db   79653 TTCTTGTTC 79645


RESULT 15
AP001977
LOCUS       AP001977    ...0069 bp    DNA              HTG      12-AUG-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-248G21 map 11q, WORKING DRAFT
            SEQUENCE, 15 unordered pieces.
ACCESSION   AP001977
VERSION     AP001977.   ...:.9798618
KEYWORDS    HTG; HTC; HASE1; HTGS_DRAFT.
SOURCE      Homo sa...   . DNA, clone:RP11-248G21.
ORGANISM    Homo sa...
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
     misc_feature   /clone_lib="RPCI-11.1"
                    1..81014
                    /note="assembly_fragment:00789
                    fragment_chain:1"
     misc_feature   81115..105285
                    /note="assembly_fragment:04577
                    fragment_chain:1"
     misc_feature   105386..119420
                    /note="assembly_fragment:00534"
     misc_feature   119521..141236
                    /note="assembly_fragment:00652"
     misc_feature   141337..180096
                    /note="assembly_fragment:03289"
     misc_feature   180197..182529
                    /note="assembly_fragment:05974"
BASE COUNT     49135 a  43556 c  42664 g  46674 t    500 others
ORIGIN


Query Match                     7.8%;     Score 35.8;   DB 82;   Length 182529;
Best Local Similarity   52.3%;     Pred. No. 7.6;
Matches    79;   Conservative     0;   Mismatches     72;   Indels     0;   Gaps     0;

Qy     214 ggccctgatcggatttcttaccattgtcggattaccgctatggcgcttcatcaaca 273
           || ||| ||| ||| | |||| ||||| || |||||| |||||| ||  ||  ||
Db   136425 GGCTCTGCCCGGCTAGCACTTAGGAAAGAAGGATGTATGCCCCTGGGCGTGCTGGGAGCACT 136366
Qy     274 gaggatgccctagtatcaccctacccacccccgtgaccaacctggccttaaataaa 333
           | ||| | |||| |||  ||||| | |||||||| |||| |||||| |||  | |
Db   136365 GTGGGAACTCTGGGATCCCCCTCCTTAAGGCGTGTCTACCCTGATTTGGAAACAATA 136306
Qy     334 cctatcttgtcattcggaataaacttcc 364
           | | ||||| || | |||| ||| | ||
Db   136305 CAGAGGCTTCTCTCCTGCAAGAAACATGCC 136275


RESULT 14
AC073897/c
LOCUS       AC073897    180538 bp    DNA              PRI      10-APR-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-718O11 map 15q21.1, complete
            sequence.
ACCESSION   AC073897
VERSION     AC073897.6    GI:13569968
KEYWORDS    HTG.
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180538)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Sequencing of human chromosome 15 D15S146-D15S117 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180538)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood P., Harrison,G., Kaur,A.,
            Madan,A., Nesbitt,R., Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-2000) Multimegabase Sequencing Center, Institute
            for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
            98105, USA
REFERENCE   3  (bases 1 to 180538)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2001) Multimegabase Sequencing Center, Institute
            for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
            98105, USA
  COMMENT   On Apr 10, 2001 this sequence version replaced gi:13173610.
            --------- Genome Center
            Center: Multimegabase Sequencing Center
```

```
REFERENCE   1  (bases 1 to 160069)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Homo sapiens 160,069 genomic DNA of 11q
JOURNAL     Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 160069)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (02-MAY-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN) Genomic Sciences Center (GSC);
            1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8117592.
            ------------- Genome Center
              Center: RIKEN Genomic Sciences Center(GSC)
              Center code: RIKEN
              Web site: http://hgp.gsc.riken.go.jp/
              Contact: hattori@gsc.riken.go.jp
            ------------- Project Information
              Center project name: HumDraft11
              Center clone name: RP11-248G21
            ------------- Summary Statistics
              Sequencing vector: PCR products; 100% of reads
              Chemistry: Dye-terminator ET-amersham; 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 155174 bases at least Q40
              Consensus quality: 156927 bases at least Q30
              Consensus quality: 157970 bases at least Q20
              Insert size: 158669; sum-of-contigs
              Quality coverage: 10.38x in Q20 bases; sum-of-contigs
            -------------
            NOTE: This is a 'working draft' sequence. It currently consists of
            15 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs of N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved.
                 1     26673 contig of 26673 bp in length
             26774     45694 contig of 18921 bp in length
             45795     64863 contig of 19069 bp in length
             64964     80530 contig of 15567 bp in length
             80631     93516 contig of 12886 bp in length
             93617    106718 contig of 13102 bp in length
            106819    120413 contig of 13595 bp in length
            120514    131809 contig of 11296 bp in length
            131910    139135 contig of 7226 bp in length
            139236    145638 contig of 6403 bp in length
            145739    150381 contig of 4643 bp in length
            150482    154739 contig of 4258 bp in length
            154840    157726 contig of 2887 bp in length
            157827    158335 contig of 509 bp in length
            158436    160069 contig of 1634 bp in length.
            NOTE: This is a 'working draft' sequence. It currently
            consists of 15 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
*                 1   26673: contig of 26673 bp in length
*   26674     26773: gap of    100 bp
*   26774     45694: contig of 18921 bp in length
*   45695     45794: gap of    100 bp
*   45795     64863: contig of 19069 bp in length
*   64864     64963: gap of    100 bp
*   64964     80530: contig of 15567 bp in length
*   80531     80630: gap of    100 bp
*   80631     93516: contig of 12886 bp in length
*   93517     93616: gap of    100 bp
*   93617    106718: contig of 13102 bp in length
*  106719    106818: gap of    100 bp
*  106819    120413: contig of 13595 bp in length
*  120414    120513: gap of    100 bp
*  120514    131809: contig of 11296 bp in length
*  131810    131909: gap of    100 bp
*  131910    139135: contig of 7226 bp in length
*  139136    139235: gap of    100 bp
*  139236    145638: contig of 6403 bp in length
*  145639    145738: gap of    100 bp
*  145739    150381: contig of 4643 bp in length
*  150382    150481: gap of    100 bp
*  150482    154739: contig of 4258 bp in length
*  154740    154839: gap of    100 bp
*  154840    157726: contig of 2887 bp in length
*  157727    157826: gap of    100 bp
*  157827    158335: contig of 509 bp in length
*  158336    158435: gap of    100 bp
*  158436    160069: contig of 1634 bp in length.
FEATURES             Location/Qualifiers
     source          1..160069
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-248G21"
     misc_feature    1..26673
                     /note="assembly_fragment"
     misc_feature    26774..45694
                     /note="assembly_fragment"
     misc_feature    45795..64863
                     /note="assembly_fragment"
     misc_feature    64964..80530
                     /note="assembly_fragment"
     misc_feature    80631..93516
                     /note="assembly_fragment"
     misc_feature    93617..106718
                     /note="assembly_fragment"
     misc_feature    106819..120413
                     /note="assembly_fragment"
     misc_feature    120514..131809
                     /note="assembly_fragment"
     misc_feature    131910..139135
                     /note="assembly_fragment"
     misc_feature    139236..145638
                     /note="assembly_fragment clone_end:SP6 vector_side:right"
     misc_feature    145739..150381
                     /note="assembly_fragment"
     misc_feature    150482..154739
                     /note="assembly_fragment"
     misc_feature    154840..157726
                     /note="assembly_fragment"
     misc_feature    157827..158335
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    158436..160069
                     /note="assembly_fragment"

BASE COUNT     48004 a    30901 c    31466 g    48298 t    1400 others
ORIGIN

Query Match             7.5%;   Score 35;   DB 83;   Length 160069;
Best Local Similarity   56.0%;   Pred. No. 13;
Matches  65;  Conservative  0;  Mismatches  51;  Indels  0;  Gaps  0;

Qy    317 ttggcctttaaataaaacctatctttgtcattcggaataaactttcctgtctttcagca 376
          ||||||| ||||||| |||||| |||||| ||| || |  ||| |  ||||| | ||| |||
Db  18067 TTCTCTTCTTGTAATATCTATTCTTGACATTACGGTGAAGATGTTTTCTCTTAAAACA 18126

Qy    377 aatgaagaacccctgctcngtccttcgggcgggccttttgcacgcttttggacct 432
          |||||| |  |||| |||| ||| | ||| |||| ||| | ||| | ||||| |||
Db  18127 ACTGAAAATTCTCTGCACTTTGTTCTGCCTGGATTTGAGTACCCTGTCATTTCTT 18182
```

Wed Nov. 7 10:16:51 2001

Search completed: November 7, 2001, 04:14:09
Job time: 2979 sec

OM nucleic - nucleic search, using sw model

        GenCore version 4.5
    Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:       November  7, 2001, 03:25:40 ; Search time 130.19 Seconds
                                 (without alignments)
                              2242.678 Million cell updates/sec

Title:              US-09-521-640-2
Perfect score:      465
Sequence:           1 gggagtacgtgagttcac.......cctctcttgttcaaaacacn 465

Scoring table:      IDENTITY_NUC
                    Gapop 10.0 , Gapext 1.0

Searched:           730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters:    1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:    Minimum Match 0%
                    Maximum Match 100%
                    Listing first 45 summaries

Database :          N_Geneseq_0601:*
    1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
    2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
    3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
    4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
    5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
    6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
    7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
    8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
    9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
   10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
   11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
   12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
   13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
   14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
   15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
   16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
   17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
   18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
   19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
   20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
   21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
   22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

                    SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

| Result No. | Score | % Query Match | Score Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 34 | 7.3 | 549 | 21 | AAC95226 | Cat flea head and |
| 2 | 32.4 | 7.0 | 607 | 21 | AAC68823 | Human head/neck tu |
| c 3 | 31.8 | 6.8 | 89047 | 22 | AAF28547 | Genomic fragment # |
| 4 | 31.6 | 6.8 | 963 | 21 | AAA98219 | Human retrovirus L |
| 5 | 30.6 | 6.6 | 50000 | 21 | AAA96367 | Polymorphic repeat |
| 6 | 30.2 | 6.5 | 1166 | 20 | AAZ19440 | M. tuberculosis an |
| c 7 | 30.2 | 6.5 | 1166 | 21 | AAZ19228 | M. tuberculosis re |
| c 8 | 30 | 6.5 | 684 | 21 | AAC52130 | Arabidopsis thalia |
| c 9 | 30 | 6.5 | 757 | 21 | AAA02266 | Human colon cancer |
| c 10 | 30 | 6.5 | 1233 | 21 | AAC38681 | Arabidopsis thalia |
| c 11 | 30 | 6.5 | 1329 | 21 | AAC34844 | Arabidopsis thalia |

| | | | | | | | | Description |
|---|---|---|---|---|---|---|---|---|
| c | 12 | 30 | 6.5 | 5616 | 22 | AAF63962 | | Human tankyrase1 3 |
| | 13 | 29.6 | 6.4 | 798 | 21 | AAC53576 | | Arabidopsis thalia |
| | 14 | 29.6 | 6.4 | 1095 | 21 | AAC77435 | | Human ORFX ORF2990 |
| | 15 | 29.4 | 6.3 | 640 | 21 | AAC95117 | | Cat flea hindgut a |
| | 16 | 29.4 | 6.3 | 743 | 21 | AAA01800 | | Human colon cancer |
| | 17 | 29.4 | 6.3 | 1053 | 20 | AAX86032 | | Nucleic acid encod |
| c | 18 | 29.4 | 6.3 | 1053 | 20 | AAX86033 | | Membrane penetrati |
| | 19 | 29.4 | 6.3 | 1475 | 19 | AAV43795 | | Rodent chemokine r |
| | 20 | 29.4 | 6.3 | 9289 | 20 | AAX34655 | | Hexaploid wheat DB |
| c | 21 | 29.4 | 6.3 | 910715 | 20 | AAX20248 | | Borrelia burgdorfe |
| | 22 | 28.8 | 6.2 | 477 | 18 | AAX83232 | | Breast cancer tumo |
| c | 23 | 28.8 | 6.2 | 477 | 19 | AAV68830 | | DNA molecule encod |
| | 24 | 28.8 | 6.2 | 477 | 21 | AAC80755 | | Human breast tumou |
| c | 25 | 28.6 | 6.2 | 383 | 19 | AAV61979 | | Human mu-opioid re |
| | 26 | 28.6 | 6.2 | 1116 | 20 | AA216998 | | Human gene express |
| | 27 | 28.6 | 6.2 | 3336 | 14 | AAQ64652 | | Human Mannose-Bind |
| | 28 | 28.6 | 6.2 | 3592 | 14 | AAQ53529 | | Human Mannose-Bind |
| | 29 | 28.6 | 6.2 | 3605 | 20 | AAZ07143 | | Human mannan-bind1 |
| c | 30 | 28.4 | 6.1 | 403 | 22 | AAF66785 | | Novel human polynu |
| | 31 | 28.4 | 6.1 | 646 | 21 | AAC76174 | | Human ORFX ORF1729 |
| c | 32 | 28.4 | 6.1 | 3444 | 16 | AAT05249 | | CryIF/cryIA(b) chi |
| c | 33 | 28.4 | 6.1 | 3444 | 16 | AAT05251 | | CryIF/cryIA(b) chi |
| c | 34 | 28.4 | 6.1 | 3444 | 17 | AAT18701 | | CryIF/CryIA(b) chi |
| c | 35 | 28.4 | 6.1 | 3444 | 17 | AAT18723 | | CryIF/CryIA(b) cod |
| c | 36 | 28.4 | 6.1 | 3444 | 19 | AAV62080 | | Plasmid pMYC2244 c |
| c | 37 | 28.4 | 6.1 | 3444 | 19 | AAV62082 | | Plasmid pMYC2523 c |
| c | 38 | 28.4 | 6.1 | 3450 | 16 | AAT05269 | | CryIA(c)/cryIF/cry |
| c | 39 | 28.4 | 6.1 | 3450 | 17 | AAT18721 | | CryIA(c)/CryIF/Cry |
| c | 40 | 28.4 | 6.1 | 3450 | 19 | AAV62079 | | Plasmid pMYC2239 B |
| c | 41 | 28.4 | 6.1 | 3522 | 12 | AAQ10182 | | Lepidopteran-activ |
| c | 42 | 28.4 | 6.1 | 3522 | 14 | AAQ47291 | | Delta endotoxin ge |
| c | 43 | 28.4 | 6.1 | 3522 | 16 | AAT05270 | | CryIF toxin with l |
| c | 44 | 28.4 | 6.1 | 3522 | 16 | AAT05250 | | CryIF/436 chimeric |
| c | 45 | 28.4 | 6.1 | 3522 | 17 | AAT18702 | | CryIF/436 chimeric |

                    ALIGNMENTS

RESULT  1
AAC95226
ID   AAC95226 standard; cDNA; 549 BP.
XX
AC   AAC95226;
XX
DT   19-FEB-2001  (first entry)
XX
DE   Cat flea head and nerve cord (HNC) cDNA, SEQ ID NO:1721.
XX
KW   Cat flea; head and nerve cord nucleic acid; HNC; flea infestation;
KW   vaccine; antiparasitic; therapeutic target; diagnosis; detection; ss.
XX
OS   Ctenocephalides felis.
XX
PN   WO200061621-A2.
XX
PD   19-OCT-2000.
XX
PF   07-APR-2000; 2000WO-US09437.
XX
PR   09-APR-1999;  99US-0128704.
XX
PA   (HESK-) HESKA ( . ).
XX
PI   Brandt KS, Gai . ; PJ, Stinchcomb DT, Wisnewski N;
XX
DR   WPI; 2000-6565;
XX
PT   Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT   acids useful for the prevention, diagnosis and treatment of flea
PT   infestations -
XX
PS   Claim 26; Pag.  164pp; English.

```
XX
CC   The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC   acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC   or head and nerve cord (HNC) tissue. The invention also relates to the
CC   encoded proteins. The invention additionally encompasses expression
CC   constructs, recombinant viruses and recombinant cells comprising the
CC   nucleic acids of the invention, recombinant production of the proteins,
CC   the proteins, and compositions comprising the inhibitors of
CC   antibodies against the proteins, a method of identifying inhibitors of
CC   administration to an animal. The nucleic acids, and the proteins they
CC   encode may be used in the prevention, treatment and diagnosis of diseases
CC   associated with flea infestations. For example, the nucleic acids may be
CC   used to produce an HMT or HNC protein according to standard recombinant
CC   DNA methodology by inserting the nucleic acids into a host cell and
CC   culturing the cell to express the protein. The HMT and HNC nucleic acids
CC   may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC   and quantitate the presence of cat flea or other homologous nucleic acid
CC   sequences in samples. They may also be used to study the expression and
CC   function of the proteins and their role in metabolism. The HMT and HNC
CC   proteins may be used as antigens in the production of specific
CC   antibodies, and in assays to identify modulators (agonists and
CC   antagonists) of HMT and/or HNC protein expression and activity. The
CC   anti-HMT/HNC protein antibodies and antagonists may also be used to
CC   downregulate protein expression and activity. The antibodies may also be
CC   used as diagnostic agents for detecting the presence of flea polypeptides
CC   in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC   present sequence represents a cat flea HNC cDNA of the invention.
XX
SQ   Sequence 549 BP; 201 A; 102 C; 98 G; 148 T; 0 other;

Query Match            7.3%;
Best Local Similarity  88.1%;    Pred. No. 0.13;   DB 21;   Length 549;
Matches   37;  Conservative  0;  Mismatches  0;  Indels  5;  Gaps  0;

Qy   93 gccgcgcttacacgcggactgagaaacctgcgtacccgctaa 134
         |||||| |||||||||||| || ||||||||| |||||||
Db  508 gccgcgtttacacgcggactggaaacccctgcgtacccactaa 549


RESULT  2
AAC68823
ID   AAC68823 standard; cDNA; 607 BP.
XX
AC   AAC68823;
XX
DT   21-FEB-2001  (first entry)
XX
DE   Human head/neck tumour related protein partial coding sequence #20.
XX
KW   Head tumour; neck tumour; lung cancer; vaccine; cancer therapy; ss.
>
O.   Homo sapiens.
XX
PN   WO20000065053-A2.
XX
PD   02-NOV-2000.
XX
PF   21-APR-2000; 2000WO-US10687.
XX
PR   23-APR-1999;  99US-0130906.
PR   20-APR-2000;  2000US-0533870.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Dillon DC;
XX
DR   WPI; 2000-687345/67.
XX
PT   Novel polypeptides comprising immunogenic portion of head and neck
PT   tumour protein useful for treating, diagnosing and monitoring cancer
PT   such as head, neck and lung cancer -
XX
```

```
PS   Claim 3; Page 76; 77pp; English.
XX
CC   The present invention relates to a number of nucleic acid sequences which
CC   encode proteins associated with head, neck and lung tumours. These
CC   tumours are often not diagnosed until they have spread, and, of those who
CC   survive, most must endure alterations in facial and neck appearance as
CC   well as changes in speech, sight, smell, chewing, swallowing and taste
CC   perception. The coding sequences given, and the proteins they encode, can
CC   be used in the diagnosis, treatment and vaccination against cancer,
CC   particularly papillary and follicular carcinomas, papillary tumours,
CC   follicular adenoma, parathyroid hyperplasia, parotid cancer, lip cancer,
CC   squamous cell cancer of the tongue, oral tongue cancers and larynx
CC   cancer.
XX
SQ   Sequence 607 BP; 154 A; 159 C; 130 G; 141 T; 23 other;

Query Match            7.0%;     Score 32.4; DB 21;   Length 607;
Best Local Similarity  74.6%;    Pred. No. 0.48;
Matches   53;  Conservative  0;  Mismatches  17;  Indels  1;  Gaps  1;

Qy   82 cttaagtcactgccgcgcttacacgcggactgagaaacctgcgtacccgctaatcctcttc 141
         || |||||||||||||||||| |  |||||||||| ||| |||||| |||
Db  262 ctcattactggccgcgtttacacgcggactggaaacccgcgt-cccactaatcgcttn 320

Qy  142 acacatccct 152
         ||||||||| |
Db  321 acacatccctt 331


RESULT  3
AAF28547/C
ID   AAF28547 standard; DNA; 89047 BP.
XX
AC   AAF28547;
XX
DT   04-APR-2001  (first entry)
XX
DE   Genomic fragment #34.
XX
KW   Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW   bronchopulmonary; endocarditis; meningitis; ss.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200078968-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000WO-US16649.
XX
PR   18-JUN-1999;  99US-0140121.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lagace RE,  Patterson C,  Berg KL;
XX
DR   WPI; 2001-041427/05.
XX
PT   Genomic library for identifying diagnostic and therapeutic
PT   compositions, and for identifying virulence factors, regulatory
PT   elements and drug targets, comprises Moraxella catarrhalis nucleic
PT   acids -
XX
PS   Claim 1; Page 324-345; 545pp; English.
XX
CC   The present invention relates to a Moraxella catarrhalis genomic library
CC   comprising of a combination of 41 nucleic acid molecules (see
CC   AAF28514-AAF28554). The library has a number of uses described in the
CC   specification e.g. is useful for aidentifying diagnostic and therapeutic
CC   compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC   aerobic, gram-negative diplococcus, normally found among the bacterial
CC   flora of human upper airways. M. catarrhalis is known to cause acute,
```

CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
SQ Sequence 89047 BP; 26501 A; 17338 C; 19060 G; 26147 T; 1 other;

Query Match 6.8%; Score 31.8; DB 22; Length 89047;
Best Local Similarity 50.3%; Pred. No. 8.1;
Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 119 cctgcgtaccgcgtaatctcttcacacatccctatccgcactggcggtatctcaatgacc 178
Db 49455 CCTTAGCACCCAATGGCGTCCTTAACACCTGCACGCACCCAGCAGCGATGTGC 49396

Qy 179 accatcgcctccaaccatgccaactaatggcaatgccctgatcggattctcttacca 238
Db 395 CGCAGTGCGTGCAAACAGTCCAAGTTTTCAAGTGTGTCACCGAACCGATACCACGCTTAG 49336

Qy 239 ttgtcggatttaccgctatgcgcttcatcaaca 273
Db 49335 GTGTATTAATCACTCGCAAAAATGCCCGCATCATCA 49301

RESULT 4
AAA98219
ID AAA98219 standard; DNA; 963 BP.
XX
AC AAA98219;
XX
DT 30-JAN-2001 (first entry)
XX
DE Human retrovirus LTR DNA fragment L49-LTR = L20-LTR.
XX
KW Cell-specific expression; tissue-specific expression; gene therapy; LTR;
KW U3-R segment; long terminal repeat; retroviral expression vector; ds.
XX
OS Human retrovirus.
XX
PN WO200053789-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-EP02064.
XY
PI .0-MAR-1999; 99DE-1010650.
XX
PA (GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEI.
XX
PI Leib-Moesch C, Schoen U, Baust C;
XX
DR WPI; 2000-587442/55.
XX
PT Retroviral expression vector, useful in gene therapy, contains a
PT promoter from a human endogenous retrovirus to provide cell-specific
PT expression -
XX
PS Disclosure; Page 66; 67pp; German.
XX
CC This invention describes a novel retroviral expression vector (A)
CC containing DNA sequences (I) for packaging vector RNA and for
CC cell-specific expression of proteins or peptides encoding by heterologous
CC DNA (II). The sequences controlling cell-specific expression contain a
CC cell-specifically regulatable promoter region (P) from a human endogenous
CC retrovirus (HERV) DNA sequence. The invention also describes (a) mRNA and
CC RNA of (A); (b) prokaryotic and eukaryotic cells containing (A); (c)
CC eukaryotic cells containing (A) in integrated form; (d) virions
CC containing a retroviral expression vector RNA derived from (A); (e) a
CC method for producing the virions of (d); (f) a method for incorporating
CC protein-encoding nucleic acid sequences into a eukaryotic cell by
CC infection with the virions of (d); and (g) a retroviral vector system
CC containing (A) and a packaging cell line, that contains at least one
CC (recombinant) retrovirus construct that encodes for the packaging

CC proteins of (A). (A) are used for cell- or tissue-specific expression of
CC foreign genes fc. gene therapy and to produce virions for introducing
CC (II) into the chromosomal DNA of eukaryotic cells, preferably mammalian
CC and specifically human. (A) retain the advantages of usual retroviral
CC promoters with all the signal structures required for transcription in a
CC small region within the U3-R segment, but without their disadvantages
CC (excessive strength and limited cell specificity). Since (A) are derived
CC from endogenous (harmless) viral sequences, they do not introduce any new
CC viral sequences into the genome and recombination will not create new
CC types of retrovirus. The promoters provide cell or tissue specific
CC expression, according to which HERV they are derived from.
XX
SQ Sequence 963 BP; 241 A; 223 C; 218 G; 281 T; 0 other;

Query Match 6.8%; Score 31.6; DB 21; Length 963;
Best Local Similarity 56.3%; Pred. No. 1.1;
Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 303 cccgtgaccaacctggccttaaataaaacctatcttgtcatttcggaataaacttt 362
Db 848 cccggcccagctattgtctcttatcctccttgtcttgtctttattattacaatct 907

Qy 363 cctgtcttcagcaaatgaagaacccctgctcngtcccttggg 405
Db 908 ctgtctctgcacacagggagaacctgctaagcccgtagg 950

RESULT 5
AAA96367
ID AAA96367 standard; DNA; 50000 BP.
XX
AC AAA96367;
XX
DT 08-FEB-2001 (first entry)
XX
DE Polymorphic repeat microsatellite sequences present in the CTLA4 locus.
XX
KW Autoimmune disease; polymorphic microsatellite repeat; PMR; CD28 gene;
KW ICOS gene; CTLA4 gene; costimulatory receptor gene locus; CGRL; lupus;
KW insulin-dependent diabetes mellitus; IDDM; Addison's disease; leprosy;
KW Graves disease; autoimmune hypothyroidism; myasthenia gravis; thymoma;
KW thyroiditis; postpartum thyroiditis; rheumatoid arthritis;
KW Hashimoto's disease; coeliac disease; ss.
XX
OS Homo sapiens.
XX
PN WO200056856-A2.
XX
PD 28-SEP-2000.
XX
PF 24-MAR-2000; 2000WO-US07938.
XX
PR 25-MAR-1999; 99US-0126215.
XX
PA (GEMY ) GENETICS INST INC.
XX
PI Ling V, Wu P, Gray GS;
XX
DR WPI; 2000-628257/60.
XX
PT Determining predisposition of humans to develop autoimmune disease
PT involves detecting polymorphic microsatellite repeat sequence within
PT human costimulatory receptor gene locus -
XX
PS Disclosure; Page 128-142; 160pp; English.
XX
CC Two human bacterial artificial chromosome (BAC) clones that included
CC and flanked the human CTLA-4 locus were cloned and sequenced The
CC sequence data was assembled into a contiguous sequence that is presented
CC in AAA96363-68 AAA96363-64 comprise BAC clone 22700, and AAA96365-68
CC comprise BAC clone 22608. The sequences contain polymorphic
CC microsatellite repeat (PMR) sequences. The specification describes a

```
CC.   method for determining the predisposition of a human subject to develop
CC    autoimmune disease. The method comprises detecting a PMR sequence in the
CC    CD28, ICOS gene or CTLA4 gene of the human costimulatory receptor gene
CC    locus (hCRGL). PMR sequences vary in length among individuals and can be
CC    amplified to generate products that differ in size. These products can
CC    then be detected by rapid and convenient high resolution processes. The
CC    method is useful for determining the predisposition of insulin-dependent
CC    diabetes mellitus (IDDM), Addison's disease, Graves disease, autoimmune
CC    hypothyroidism, myasthenia gravis, thymoma, lupus, thyroiditis,
CC    postpartum thyroiditis, rheumatoid arthritis, Hashimoto's disease,
CC    coeliac disease and leprosy. PMR sequences within hCRGL are useful as
CC    markers in a variety of assays and in the field of forensic medicine,
CC    disease diagnosis and human genome mapping.
XX
SQ    Sequence 50000 BP; 14612 A; 9948 C; 10072 G; 15368 T; 0 other;
```

```
Query Match             6.6%;  Score 30.6; DB 21; Length 50000;
Best Local Similarity  45.3%;  Pred. No. 16;
  Matches  111; Conservative    0; Mismatches  134; Indels    0; Gaps     0;

Qy      14 agttcactcaatattagccctcagactcgctgtaagacctcggtccagccgaatcccct 73
           ||| || |||| |||| ||| |  | | || ||  ||| |  ||||| || | |
Db    5472 attcctttcagcttcaagctacccctccccaggagttccaggaccacacaaaagcagg 26531

Qy      74 tatgagcgcttaagtcactgccgcgctacacggcgdactgagaaacctgcgtaccgcta 133
           || | |||||| ||||  | |||  | ||| ||||| |  ||  ::  ::  || ||
Db   26532 tctcttgcctcagtccatgtcatttatgttcttagttctgaatcattcatctgacatatctctt 26591

Qy     134 atcctttcacacatccccatcgcactgcgggtatctcaatgaccacatcgccttcaa 193
           |||| |||| || || |||  | |||| || |||||||| |||||||| |  |  |
Db   26592 atcacctccaactcctacaggaatctggcagtaccacatgcccaccatgatcattgga 26651

Qy     194 ccatgcaactaatggcaatggcctgatcggatttctcttaccattgtcggatttaccc 253
           ||  |  |||  ||| || | ||| |  ||||||||| ||||| || |||| |||||
Db   26652 tgcacttgctatatgatatggttggctgtgttcccaccccaaatcctcatgttgaatt 26711

Qy     254 gctat 258
           || ||
Db   26712 gcaat 26716
```

```
RESULT  6
AA219440
ID    AA219440 standard; cDNA; 1166 BP.
XX
AC    AA219440;
XX
DT    05-NOV-1999  (first entry)
XX
DE    M. tuberculosis antigen 3' Erdsn-7 cDNA sequence.
XX
KW    .;
KW    Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW    immunotherapy; diagnosis; immunisation; vaccine; infection;
KW    immune response; skin test; ss.
XX
OS    Mycobacterium tuberculosis.
XX
PN    WO942076-A2.
XX
PD    26-AUG-1999.
XX
PF    17-FEB-1999;  99WO-US03268.
XX
PR    05-MAY-1998;  98US-0072967.
PR    18-FEB-1998;  98US-0025197.
XX
PA    (CORI-) CORIXA CORP.
XX
PI    Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI    Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR    WPI; 1999-527409/44.
```

```
XX
PT    New antigens from Mycobacterium tuberculosis useful in diagnostic
PT    skin tests and protective or therapeutic vaccines or compositions
PS    Claim 4; Page 259; 299pp; English.
XX
CC    The present invention describes polypeptides comprising an immunogenic
CC    part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC    are vaccines and fusion protein containing M. tuberculosis Ag's.
CC    M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC    other polypeptides fragments, can be used in pharmaceutical compositions
CC    or vaccines to generate a protective or therapeutic immune response to
CC    M. tuberculosis and as reagents in skin tests for diagnosis of
CC    tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC    by, T, B or natural killer cells and/or macrophages in
CC    tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
CC    AAY39225 are used in the exemplification of the present invention.
XX
SQ    Sequence 1166 BP; 117 A; 371 C; 121 G; 207 T; 350 other;
```

```
Query Match             6.5%;   Score 30.2; DB 20; Length 1166;
Best Local Similarity  28.2%;  Pred. No. 3.6;
  Matches   71; Conservative   46; Mismatches  135; Indels    0; Gaps     0;

Qy      52 cctcgtccagccgaatccctttatgtagcgcttaagtcactgccgcgcttacacgcggac 111
           ||:: ||| :||: :|:   | || |||||| |: |  ||  :: :|| ||||| |
Db     797 cckmtcctcctmckccymwntccnmkyncctccnmtcmtckytcctcncnmrycyyyak 856

Qy     112 tgagaaacctgctaccgctaatcctccacacatccccatatccctgcggtatctc 171
           | | |  :| : :| |:  : :|  :||:|   | : :|| : :|  |||
Db     857 cakcnmctcccankmcakctkctcccccakmksacnckcccwccctcatccwctctc 916

Qy     172 aatgaccacatcgcctccaaccatgccaactaatggcaatggccctgatcggatttct 231
           :| ||   |||:    :| :: : :||     |  ||  ||  ||| |  | |:
Db     917 wctyatctkctcwcnycymymkmcacnckcyaytcnactmmwncancnctctcnyct 976

Qy     232 cttaccattgtcggatttaccgcttatggcgcttcatcaacagaggatgcctagtatca 291
           |: :: || :: : ||:||  || : : ::|  | ||   :|| :|| | || |
Db     977 cwckacgtyckcckctmckcnymcnrwctyrcctckkccncccrnckmcmkctmctctcc 1036

Qy     292 ccctaccacac 303
           : |||:|:|
Db    1037 wmkmtcccwccc 1048
```

```
RESULT  7
AA219228
ID    AA219228 standard; cDNA; 1166 BP.
XX
AC    AA219228;
XX
DT    05-NOV-1999  (first entry)
XX
DE    M. tuberculosis recombinant antigen cDNA encoding 3' Erdsn-8.
XX
KW    Antigen; diagnosis; detection; infection; antibody; immunisation;
KW    vaccine; immunity; ss.
XX
OS    Mycobacterium tuberculosis.
XX
PN    WO942118-A2.
XX
PD    26-AUG-1999.
XX
PF    17-FEB-1999;  99WO-US03265.
XX
PR    05-MAY-1998;   98US-0072596.
PR    18-FEB-1998;   98US-0024753.
XX
PA    (CORI-) CORIXA CORP.
XX
PI    Campos-Neto A,  Dillon DC,  Hendrickson RC,  Houghton R;
```

```
PI   Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR   WPI; 1999-527416/44.
XX
PT   New polypeptide comprising antigenic portions of M. tuberculosis
XX
PS   Claim 4; Page 304; 322pp; English.
XX
CC   This invention describes novel recombinant antigens and their encoding
CC   nucleic acids derived from Mycobacterium tuberculosis. The novel
CC   polypeptides are useful for detecting M. tuberculosis infection in a
CC   biological sample by detecting antibodies which bind with the
CC   polypeptides, and are useful as vaccines for immunizing against
CC   M. tuberculosis infection. The new detection methods are needed as
CC   current vaccination strategies do not provide 100% immunity.
XX
SQ   Sequence 1166 BP; 117 A; 371 C; 121 G; 207 T; 350 other;
```

```
Query Match              6.5%;   Score 30.2;  DB 20;  Length 1166;
Best Local Similarity   28.2%;   Pred. No. 3.6;
Matches    71;  Conservative   46;  Mismatches  135;  Indels   0;  Gaps   0;

Qy    52 cctcggtcagccgaatcccttatgagcgctaagtcactgccgcgcttacacgcggac 111
         ||::|||:||::::||::||:|:||::::|:::||::|::|::::::|:::::|-|--|
Db   797 cckmtctccttmckccymwcntccmkynccctccnmtcmtckytcctcncnmrycyyyak 856

Qy   112 tgagaaacctggtaccgctaatctcttcacacatccctatcgcactggcggtatctc 171
         ::||:|||:|||:::||:|:||:::||:|:::||:::|:::::::||||
Db   857 cakcnmctcccccankmcakctktctcccccakmksacnckcccwcctctatccwctctc 916

Qy   172 aatgaccacccatcgccttccaacatgccaactaatggccaatggccctgatcggatttct 231
         ::||:||:|||:||:|:::::||:|    :    :|||    :|| :   |||:||
Db   917 wctyatcctckctcwcnycmymkmcacnckcyaytcnactmnmwnccancnctctctnyct 976

Qy   232 cttaccattgtcggattacccgctatggcgcttcatcaacagaggatgccctagtatca 291
         |:   |-|:::|:|:|||   :   :||   ::||:|::    |::||||
Db   977 cwckacgtyckcckctmckccnymcnrwctyrcctckkccnccrnckmckctmctctcc 1036

Qy   292 cctaccacac 303
         |||:|::|-|
Db  1037 wmkmtcccwccc 1048
```

```
RESULT   8
AA'  '30/c
IN   AC52130 standard; DNA; 684 BP.
XX
AC   AAC52130;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 70438.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
```

```
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130510.
PR   23-APR-1999;  99US-0130851.
PR   28-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0132048.
PR   30-APR-1999;  99US-0132407.
PR   04-MAY-1999;  99US-0132484.
PR   05-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   06-MAY-1999;  99US-0132487.
PR   07-MAY-1999;  99US-0132863.
PR   11-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   14-MAY-1999;  99US-0134221.
PR   14-MAY-1999;  99US-0134370.
PR   18-MAY-1999;  99US-0134768.
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0135124.
PR   21-MAY-1999;  99US-0135353.
PR   24-MAY-1999;  99US-0135629.
PR   25-MAY-1999;  99US-0136021.
PR   27-MAY-1999;  99US-0136392.
PR   28-MAY-1999;  99US-0136782.
PR   01-JUN-1999;  99US-0137222.
PR   03-JUN-1999;  99US-0137528.
PR   04-JUN-1999;  99US-0137502.
PR   07-JUN-1999;  99US-0137724.
PR   08-JUN-1999;  99US-0138094.
PR   10-JUN-1999;  99US-0138540.
PR   10-JUN-1999;  99US-0138847.
PR   14-JUN-1999;  99US-0139119.
PR   16-JUN-1999;  99US-0139452.
PR   16-JUN-1999;  99US-0139453.
PR   17-JUN-1999;  99US-0139492.
PR   18-JUN-1999;  99US-0139454.
PR   18-JUN-1999;  99US-0139455.
PR   18-JUN-1999;  99US-0139456.
PR   18-JUN-1999;  99US-0139457.
PR   18-JUN-1999;  99US-0139458.
PR   18-JUN-1999;  99US-0139459.
PR   18-JUN-1999;  99US-0139460.
PR   18-JUN-1999;  99US-0139461.
PR   18-JUN-1999;  99US-0139462.
PR   18-JUN-1999;  99US-0139463.
PR   18-JUN-1999;  99US-0139750.
PR   18-JUN-1999;  99US-0139763.
PR   21-JUN-1999;  99US-0139817.
PR   22-JUN-1999;  99US-0139899.
PR   23-JUN-1999;  99US-0140353.
PR   23-JUN-1999;  99US-0140354.
PR   24-JUN-1999;  99US-0140695.
PR   28-JUN-1999;  99US-0140823.
PR   29-JUN-1999;  99US-0140991.
PR   30-JUN-1999;  99US-0141287.
PR   01-JUL-1999;  99US-0141842.
PR   02-JUL-1999;  99US-0142154.
PR   06-JUL-1999;  99US-0142055.
PR   06-JUL-1999;  99US-0142390.
PR   08-JUL-1999;  99US-0142803.
PR   09-JUL-1999;  99US-0142920.
PR   12-JUL-1999;  99US-0142977.
PR   13-JUL-1999;  99US-0143542.
PR   14-JUL-1999;  99US-0143624.
PR   15-JUL-1999;  99US-0144005.
PR   16-JUL-1999;  99US-0144085.
PR   16-JUL-1999;  99US-0144086.
PR   19-JUL-1999;  99US-0144325.
PR   19-JUL-1999;  99US-0144331.
```

```
PR   19-JUL-1999; 99US-0144332.
PR   19-JUL-1999; 99US-0144333.
PR   19-JUL-1999; 99US-0144334.
PR   19-JUL-1999; 99US-0144335.
PR   20-JUL-1999; 99US-0144352.
PR   20-JUL-1999; 99US-0144632.
PR   20-JUL-1999; 99US-0144884.
PR   21-JUL-1999; 99US-0144814.
PR   21-JUL-1999; 99US-0145086.
PR   21-JUL-1999; 99US-0145088.
PR   22-JUL-1999; 99US-0145085.
PR   22-JUL-1999; 99US-0145087.
PR   22-JUL-1999; 99US-0145089.
PR   22-JUL-1999; 99US-0145192.
PR   23-JUL-1999; 99US-0145145.
PR   23-JUL-1999; 99US-0145218.
PR   23-JUL-1999; 99US-0145224.
PR   26-JUL-1999; 99US-0145276.
PR   27-JUL-1999; 99US-0145913.
PR   27-JUL-1999; 99US-0145918.
PR   27-JUL-1999; 99US-0145919.
PR   28-JUL-1999; 99US-0145951.
P/   12-AUG-1999; 99US-0146386.
Pk   02-AUG-1999; 99US-0146388.
PR   02-AUG-1999; 99US-0146389.
PR   03-AUG-1999; 99US-0147038.
PR   04-AUG-1999; 99US-0147204.
PR   04-AUG-1999; 99US-0147302.
PR   05-AUG-1999; 99US-0147192.
PR   05-AUG-1999; 99US-0147260.
PR   06-AUG-1999; 99US-0147303.
PR   06-AUG-1999; 99US-0147416.
PR   09-AUG-1999; 99US-0147493.
PR   09-AUG-1999; 99US-0147935.
PR   10-AUG-1999; 99US-0148171.
PR   11-AUG-1999; 99US-0148319.
PR   12-AUG-1999; 99US-0148341.
PR   13-AUG-1999; 99US-0148565.
PR   13-AUG-1999; 99US-0148684.
PR   16-AUG-1999; 99US-0149368.
PR   17-AUG-1999; 99US-0149175.
PR   18-AUG-1999; 99US-0149426.
PR   20-AUG-1999; 99US-0149722.
PR   20-AUG-1999; 99US-0149723.
PR   20-AUG-1999; 99US-0149929.
PR   23-AUG-1999; 99US-0149902.
PR   23-AUG-1999; 99US-0149930.
PR   25-AUG-1999; 99US-0150566.
PR   26-AUG-1999; 99US-0150884.
PR   27-AUG-1999; 99US-0151065.
E    27-AUG-1999; 99US-0151066.
P.   27-AUG-1999; 99US-0151080.
PR   30-AUG-1999; 99US-0151303.
PR   31-AUG-1999; 99US-0151438.
PR   01-SEP-1999; 99US-0151930.
PR   07-SEP-1999; 99US-0152363.
PR   10-SEP-1999; 99US-0153070.
PR   13-SEP-1999; 99US-0153758.
PR   15-SEP-1999; 99US-0154018.
PR   16-SEP-1999; 99US-0154039.
PR   20-SEP-1999; 99US-0154779.
PR   22-SEP-1999; 99US-0155139.
PR   23-SEP-1999; 99US-0155486.
PR   24-SEP-1999; 99US-0155659.
PR   28-SEP-1999; 99US-0156458.
PR   29-SEP-1999; 99US-0156596.
PR   04-OCT-1999; 99US-0157117.
PR   05-OCT-1999; 99US-0157753.
PR   06-OCT-1999; 99US-0157865.
PR   07-OCT-1999; 99US-0158029.
PR   08-OCT-1999; 99US-0158232.
PR   12-OCT-1999; 99US-0158369.
PR   13-OCT-1999; 99US-0159293.

PR   13-OCT-1999; 99US-0159294.
PR   13-OCT-1999; 99US-0159295.
PR   14-OCT-1999; 99US-0159329.
PR   14-OCT-1999; 99US-0159330.
PR   14-OCT-1999; 99US-0159331.
PR   14-OCT-1999; 99US-0159637.
PR   14-OCT-1999; 99US-0159638.
PR   18-OCT-1999; 99US-0159584.
PR   21-OCT-1999; 99US-0160741.
PR   21-OCT-1999; 99US-0160767.
PR   21-OCT-1999; 99US-0160768.
PR   21-OCT-1999; 99US-0160770.
PR   21-OCT-1999; 99US-0160814.
PR   21-OCT-1999; 99US-0160815.
PR   22-OCT-1999; 99US-0160980.
PR   22-OCT-1999; 99US-0160981.
PR   22-OCT-1999; 99US-0160989.
PR   25-OCT-1999; 99US-0161404.
PR   25-OCT-1999; 99US-0161405.
PR   25-OCT-1999; 99US-0161406.
PR   26-OCT-1999; 99US-0161159.
PR   26-OCT-1999; 99US-0161360.
PR   26-OCT-1999; 99US-0161361.
PR   28-OCT-1999; 99US-0161920.
PR   28-OCT-1999; 99US-0161992.
PR   28-OCT-1999; 99US-0161993.
PR   29-OCT-1999; 99US-0162142.


Query Match             6.5%;    Score 30; DB 21; Length 684;
Best Local Similarity  53.4%;              Pred. No. 3.3;
  Matches  63; Conservative  0; Mismatches 55; Indels  0; Gaps  0;

Qy    165 gtatctcaatgaccaccatgcgcttccaacatgccaactaatggccctgatcg 224
          ||  ||  |||||||| ||  |||| |||  || | | | |  | |||| ||||
Db    169 GTTTCCGATCGACCACCTCGACTTCGATCCAACCATACCTCTCCCTATACATGATCT 110

Qy    225 gattctcttccattgtcggatttaccgcgctatggcgcttcatcaacagaggatgcc 282
          | |||||  |  ||  | | ||||  | |||  ||  |  | | |  |||
Db    109 AAAATCTCTCCGAAATTCCGAATCTCCCCTCCACCGGAGATCCACGCGCGGTGAAAGCC 52


RESULT   9
AAA02266
ID   AAA02266 standard; cDNA; 757 BP.
XX
AC   AAA02266;
XX
DT   19-MAY-2000   (first entry)
XX
DE   Human colon cancer cell line polynucleotide sequence SEQ ID NO:2257.
XX
KW   Human; colon cancer; tumour; diagnosis; gene expression product;
KW   probe; detection; cancerous state; metastasis; identification;
KW   breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW   oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9958675-A2.
XX
PD   18-NOV-1999.
XX
PF   13-MAY-1999; 99WO-US10602.
XX
PR   14-MAY-1998; 98US-0085426.
PR   15-MAY-1998; 98US-0085537.
PR   15-MAY-1998; 98US-0085696.
PR   21-OCT-1998; 98US-0105234.
PR   27-OCT-1998; 98US-0105877.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
```

PI  Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI  Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI  Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI  Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
DR  WPI; 2000-126369/11.
XX
PT  Polynucleotide library used to determine cancerous states of mammalian
    cells -
XX
PS  Claim 1; Page 886; 1097pp; English.
XX
CC  AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC  libraries constructed from human colon cancer cell lines. The present
CC  invention also describes a method of detecting differentially expressed
CC  genes correlated with a cancerous state of a mammalian cell, comprising
CC  detecting at least one differentially expressed gene product in a test
C   sample derived from a cell suspected of being cancerous, where detection
CC  of the differentially expressed gene product is correlated with a
CC  cancerous state of the cell from which the test sample was derived.
CC  The polynucleotides sequences can be used in a method for detecting
CC  differentially expressed genes correlated with a cancerous state of a
CC  mammalian cell. The polynucleotides can also be used as probes for
CC  detecting and mapping related genes. They can be used in diagnosis and
CC  prognosis of diseases and disorders (e.g. identification of
CC  pre-metastatic or metastatic cancerous states, stages of cancer, or
CC  responsiveness of cancer to therapy). This is particularly for breast
CC  cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC  negative breast cancer, lung cancer, and colon cancer.
XX
SQ  Sequence 757 BP; 155 A; 205 C; 168 G; 199 T; 30 other;

Query Match          6.5%;       Score 30;  DB 21;   Length 757;
Best Local Similarity  68.4%;     Pred. No. 3.4;
Matches   39;   Conservative   0;   Mismatches   18;   Indels   0;   Gaps   0;

Qy  223 cggattctcttaccattgtcggatttaccgctatgcgcttcatcaacagaggat 279
        |||||||| ||||||| ||||||| ||||||||| || ||| ||| || |
Db  700 cggattctcctaccattgncggatttacacggantgggcacctcaataacaattgnt 756

RESULT 10
AAC38681/c
ID  AAC38681 standard; DNA; 1233 BP.
X'
AL  AAC38681;
XX
DT  17-OCT-2000  (first entry)
XX
DE  Arabidopsis thaliana DNA fragment SEQ ID NO: 21860.
XX
KW  Hybridisation assay; genetic mapping; gene expression control;
KW  protein identification; signal transduction pathway;
KW  metabolic pathway; promoter; termination sequence; ss.
XX
OS  Arabidopsis thaliana.
XX
PN  EP1033405-A2.
XX
PD  06-SEP-2000.
XX
PF  25-FEB-2000; 2000EP-0301439.
XX
PR  25-FEB-1999;  99US-0121825.
PR  05-MAR-1999;  99US-0123180.
PR  09-MAR-1999;  99US-0123548.
PR  23-MAR-1999;  99US-0125788.
PR  25-MAR-1999;  99US-0126264.
PR  29-MAR-1999;  99US-0126785.
PR  01-APR-1999;  99US-0127462.
PR  06-APR-1999;  99US-0128234.
PR  08-APR-1999;  99US-0128714.

PR  16-APR-1999;  99US-0129845.
PR  19-APR-1999;  99US-0130077.
PR  21-APR-1999;  99US-0130449.
PR  23-APR-1999;  99US-0130510.
PR  23-APR-1999;  99US-0130891.
PR  28-APR-1999;  99US-0131449.
PR  30-APR-1999;  99US-0132048.
PR  30-APR-1999;  99US-0132407.
PR  04-MAY-1999;  99US-0132484.
PR  05-MAY-1999;  99US-0132485.
PR  06-MAY-1999;  99US-0132486.
PR  06-MAY-1999;  99US-0132487.
PR  07-MAY-1999;  99US-0132863.
PR  11-MAY-1999;  99US-0134256.
PR  14-MAY-1999;  99US-0134218.
PR  14-MAY-1999;  99US-0134219.
PR  14-MAY-1999;  99US-0134221.
PR  14-MAY-1999;  99US-0134370.
PR  18-MAY-1999;  99US-0134768.
PR  19-MAY-1999;  99US-0134941.
PR  20-MAY-1999;  99US-0135124.
PR  21-MAY-1999;  99US-0135353.
PR  24-MAY-1999;  99US-0135629.
PR  25-MAY-1999;  99US-0136021.
PR  27-MAY-1999;  99US-0136392.
PR  28-MAY-1999;  99US-0136782.
PR  01-JUN-1999;  99US-0137222.
PR  03-JUN-1999;  99US-0137528.
PR  04-JUN-1999;  99US-0137502.
PR  07-JUN-1999;  99US-0137724.
PR  08-JUN-1999;  99US-0138094.
PR  10-JUN-1999;  99US-0138540.
PR  10-JUN-1999;  99US-0138847.
PR  14-JUN-1999;  99US-0139119.
PR  16-JUN-1999;  99US-0139452.
PR  16-JUN-1999;  99US-0139453.
PR  17-JUN-1999;  99US-0139492.
PR  18-JUN-1999;  99US-0139454.
PR  18-JUN-1999;  99US-0139455.
PR  18-JUN-1999;  99US-0139456.
PR  18-JUN-1999;  99US-0139457.
PR  18-JUN-1999;  99US-0139458.
PR  18-JUN-1999;  99US-0139459.
PR  18-JUN-1999;  99US-0139460.
PR  18-JUN-1999;  99US-0139461.
PR  18-JUN-1999;  99US-0139462.
PR  18-JUN-1999;  99US-0139463.
PR  18-JUN-1999;  99US-0139750.
PR  18-JUN-1999;  99US-0139763.
PR  21-JUN-1999;  99US-0139817.
PR  22-JUN-1999;  99US-0139899.
PR  23-JUN-1999;  99US-0140353.
PR  23-JUN-1999;  99US-0140354.
PR  24-JUN-1999;  99US-0140695.
PR  28-JUN-1999;  99US-0140823.
PR  29-JUN-1999;  99US-0140991.
PR  30-JUN-1999;  99US-0141287.
PR  01-JUL-1999;  99US-0141842.
PR  01-JUL-1999;  99US-0142154.
PR  02-JUL-1999;  99US-0142055.
PR  06-JUL-1999;  99US-0142390.
PR  08-JUL-1999;  99US-0142803.
PR  09-JUL-1999;  99US-0142920.
PR  12-JUL-1999;  99US-0142977.
PR  13-JUL-1999;  99US-0143542.
PR  14-JUL-1999;  99US-0143624.
PR  15-JUL-1999;  99US-0144005.
PR  16-JUL-1999;  99US-0144085.
PR  16-JUL-1999;  99US-0144086.
PR  19-JUL-1999;  99US-0144325.
PR  19-JUL-1999;  99US-0144331.
PR  19-JUL-1999;  99US-0144332.
PR  19-JUL-1999;  99US-0144333.

```
PR   19-JUL-1999;  99US-0144334.        PR   14-OCT-1999;  99US-0159329.
PR   19-JUL-1999;  99US-0144335.        PR   14-OCT-1999;  99US-0159330.
PR   20-JUL-1999;  99US-0144352.        PR   14-OCT-1999;  99US-0159331.
PR   20-JUL-1999;  99US-0144632.        PR   14-OCT-1999;  99US-0159637.
PR   20-JUL-1999;  99US-0144884.        PR   14-OCT-1999;  99US-0159638.
PR   21-JUL-1999;  99US-0144814.        PR   18-OCT-1999;  99US-0159584.
PR   21-JUL-1999;  99US-0145086.        PR   21-OCT-1999;  99US-0160741.
PR   21-JUL-1999;  99US-0145088.        PR   21-OCT-1999;  99US-0160767.
PR   22-JUL-1999;  99US-0145085.        PR   21-OCT-1999;  99US-0160768.
PR   22-JUL-1999;  99US-0145087.        PR   21-OCT-1999;  99US-0160770.
PR   22-JUL-1999;  99US-0145089.        PR   21-OCT-1999;  99US-0160814.
PR   22-JUL-1999;  99US-0145192.        PR   21-OCT-1999;  99US-0160815.
PR   23-JUL-1999;  99US-0145145.        PR   22-OCT-1999;  99US-0160980.
PR   23-JUL-1999;  99US-0145218.        PR   22-OCT-1999;  99US-0160981.
PR   23-JUL-1999;  99US-0145224.        PR   22-OCT-1999;  99US-0160989.
PR   26-JUL-1999;  99US-0145276.        PR   25-OCT-1999;  99US-0161404.
PR   27-JUL-1999;  99US-0145913.        PR   25-OCT-1999;  99US-0161405.
PR   27-JUL-1999;  99US-0145918.        PR   25-OCT-1999;  99US-0161406.
PR   27-JUL-1999;  99US-0145919.        PR   26-OCT-1999;  99US-0161359.
PR   28-JUL-1999;  99US-0145951.        PR   26-OCT-1999;  99US-0161360.
PR   02-AUG-1999;  99US-0146386.        PR   26-OCT-1999;  99US-0161361.
PR   02-AUG-1999;  99US-0146388.        PR   28-OCT-1999;  99US-0161920.
PR   02-AUG-1999;  99US-0146389.        PR   28-OCT-1999;  99US-0161992.
PR   03-AUG-1999;  99US-0147038.        PR   28-OCT-1999;  99US-0161993.
PR   04-AUG-1999;  99US-0147204.        PR   29-OCT-1999;  99US-0162142.
PR   05-AUG-1999;  99US-0147302.
PR   05-AUG-1999;  99US-0147192.
PR   06-AUG-1999;  99US-0147260.
PR   06-AUG-1999;  99US-0147303.
PR   06-AUG-1999;  99US-0147416.
PR   09-AUG-1999;  99US-0147493.
PR   09-AUG-1999;  99US-0147935.
PR   10-AUG-1999;  99US-0148171.
PR   11-AUG-1999;  99US-0148319.
PR   12-AUG-1999;  99US-0148341.
PR   13-AUG-1999;  99US-0148565.
PR   13-AUG-1999;  99US-0148684.
PR   16-AUG-1999;  99US-0149368.
PR   17-AUG-1999;  99US-0149175.
PR   18-AUG-1999;  99US-0149426.
PR   20-AUG-1999;  99US-0149722.
PR   20-AUG-1999;  99US-0149723.
PR   20-AUG-1999;  99US-0149929.
PR   23-AUG-1999;  99US-0149902.
PR   23-AUG-1999;  99US-0149930.
PR   25-AUG-1999;  99US-0150566.
PR   26-AUG-1999;  99US-0150884.
PR   27-AUG-1999;  99US-0151065.
PR   27-AUG-1999;  99US-0151066.
PR   27-AUG-1999;  99US-0151080.
PR   30-AUG-1999;  99US-0151303.
PR   31-AUG-1999;  99US-0151438.
PR   01-SEP-1999;  99US-0151930.
PR   07-SEP-1999;  99US-0152363.
PR   10-SEP-1999;  99US-0153070.
PR   13-SEP-1999;  99US-0153758.
PR   15-SEP-1999;  99US-0154018.
PR   16-SEP-1999;  99US-0154039.
PR   20-SEP-1999;  99US-0154779.
PR   22-SEP-1999;  99US-0155139.
PR   23-SEP-1999;  99US-0155486.
PR   24-SEP-1999;  99US-0155659.
PR   28-SEP-1999;  99US-0156458.
PR   29-SEP-1999;  99US-0156596.
PR   04-OCT-1999;  99US-0157117.
PR   05-OCT-1999;  99US-0157753.
PR   06-OCT-1999;  99US-0157865.
PR   07-OCT-1999;  99US-0158029.
PR   08-OCT-1999;  99US-0158232.
PR   12-OCT-1999;  99US-0158369.
PR   13-OCT-1999;  99US-0159293.
PR   13-OCT-1999;  99US-0159294.
PR   13-OCT-1999;  99US-0159295.
```

```
Query Match         6.5%;       Score 30;  DB 21;   Length 1233;
Best Local Similarity  53.4%;   Pred. No. 4.3;
Matches   63;  Conservative   0;  Mismatches  55;  Indels   0;  Gaps   0;

Qy    165 gtatctcaatgacaccatcgccttccaactaatggccaactaatggccctgatcg 224
          ||-|||-|  ||||||| |||||| || ||| || ||| ||| || || ||| ||
Db    215 GTTTCGATCGACCACCTCCGATTCGATCCAACCATACCTCTCTCTATACATGATCT 156

Qy    225 gatttcttcttaccattgtcggatttacccgtatggcgcttcatcaacagaggatgcc 282
           | |||||  |  ||| || || ||  |  ||| || |  ||| ||  | ||| || |
Db    155 AAAATCTCTCCGAAATTCCGAATCTCCCCGTCACCGGAGATCCCACGCCCGTGAAAGCC 98
```

```
RESULT 11
AAC34844/C
ID   AAC34844 standard; DNA; 1329 BP.
XX
AC   AAC34844;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 8091.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130049.
```

us-09-521-640-2_1.rng

Wed Nov  7  10:16:52  2001

| | | |
|---|---|---|
| PR | 23-APR-1999; | 99US-0130510. |
| PR | 23-APR-1999; | 99US-0130891. |
| PR | 28-APR-1999; | 99US-0131449. |
| PR | 30-APR-1999; | 99US-0132048. |
| PR | 30-APR-1999; | 99US-0132407. |
| PR | 04-MAY-1999; | 99US-0132484. |
| PR | 05-MAY-1999; | 99US-0132485. |
| PR | 06-MAY-1999; | 99US-0132486. |
| PR | 06-MAY-1999; | 99US-0132487. |
| PR | 07-MAY-1999; | 99US-0132863. |
| PR | 11-MAY-1999; | 99US-0134256. |
| PR | 14-MAY-1999; | 99US-0134218. |
| PR | 14-MAY-1999; | 99US-0134219. |
| PR | 14-MAY-1999; | 99US-0134221. |
| PR | 18-MAY-1999; | 99US-0134370. |
| PR | 18-MAY-1999; | 99US-0134768. |
| PR | 19-MAY-1999; | 99US-0134941. |
| P. | 0-MAY-1999; | 99US-0135124. |
| Ph | 21-MAY-1999; | 99US-0135353. |
| PR | 24-MAY-1999; | 99US-0135629. |
| PR | 25-MAY-1999; | 99US-0136021. |
| PR | 27-MAY-1999; | 99US-0136392. |
| PR | 28-MAY-1999; | 99US-0136782. |
| PR | 01-JUN-1999; | 99US-0137222. |
| PR | 03-JUN-1999; | 99US-0137528. |
| PR | 04-JUN-1999; | 99US-0137502. |
| PR | 07-JUN-1999; | 99US-0137724. |
| PR | 08-JUN-1999; | 99US-0138094. |
| PR | 10-JUN-1999; | 99US-0138540. |
| PR | 10-JUN-1999; | 99US-0138847. |
| PR | 14-JUN-1999; | 99US-0139119. |
| PR | 16-JUN-1999; | 99US-0139452. |
| PR | 16-JUN-1999; | 99US-0139453. |
| PR | 17-JUN-1999; | 99US-0139492. |
| PR | 18-JUN-1999; | 99US-0139454. |
| PR | 18-JUN-1999; | 99US-0139455. |
| PR | 18-JUN-1999; | 99US-0139456. |
| PR | 18-JUN-1999; | 99US-0139457. |
| PR | 18-JUN-1999; | 99US-0139458. |
| PR | 18-JUN-1999; | 99US-0139459. |
| PR | 18-JUN-1999; | 99US-0139460. |
| PR | 18-JUN-1999; | 99US-0139461. |
| PR | 18-JUN-1999; | 99US-0139462. |
| PR | 18-JUN-1999; | 99US-0139463. |
| PR | 18-JUN-1999; | 99US-0139750. |
| P. | 18-JUN-1999; | 99US-0139763. |
| P. | 21-JUN-1999; | 99US-0139817. |
| PR | 22-JUN-1999; | 99US-0139899. |
| PR | 23-JUN-1999; | 99US-0140353. |
| PR | 23-JUN-1999; | 99US-0140354. |
| PR | 24-JUN-1999; | 99US-0140695. |
| PR | 28-JUN-1999; | 99US-0140823. |
| PR | 29-JUN-1999; | 99US-0140991. |
| PR | 30-JUN-1999; | 99US-0141287. |
| PR | 01-JUL-1999; | 99US-0141842. |
| PR | 01-JUL-1999; | 99US-0142154. |
| PR | 02-JUL-1999; | 99US-0142055. |
| PR | 06-JUL-1999; | 99US-0142390. |
| PR | 08-JUL-1999; | 99US-0142803. |
| PR | 09-JUL-1999; | 99US-0142920. |
| PR | 12-JUL-1999; | 99US-0142977. |
| PR | 13-JUL-1999; | 99US-0143542. |
| PR | 14-JUL-1999; | 99US-0143624. |
| PR | 15-JUL-1999; | 99US-0144005. |
| PR | 16-JUL-1999; | 99US-0144085. |
| PR | 16-JUL-1999; | 99US-0144086. |
| PR | 19-JUL-1999; | 99US-0144325. |
| PR | 19-JUL-1999; | 99US-0144331. |
| PR | 19-JUL-1999; | 99US-0144332. |
| PR | 19-JUL-1999; | 99US-0144333. |
| PR | 19-JUL-1999; | 99US-0144334. |
| PR | 19-JUL-1999; | 99US-0144335. |
| PR | 20-JUL-1999; | 99US-0144352. |

| | | |
|---|---|---|
| PR | 20-JUL-1999; | 99US-0144632. |
| PR | 20-JUL-1999; | 99US-0144884. |
| PR | 21-JUL-1999; | 99US-0144814. |
| PR | 21-JUL-1999; | 99US-0145086. |
| PR | 22-JUL-1999; | 99US-0145088. |
| PR | 22-JUL-1999; | 99US-0145085. |
| PR | 22-JUL-1999; | 99US-0145087. |
| PR | 22-JUL-1999; | 99US-0145089. |
| PR | 23-JUL-1999; | 99US-0145192. |
| PR | 23-JUL-1999; | 99US-0145145. |
| PR | 23-JUL-1999; | 99US-0145218. |
| PR | 26-JUL-1999; | 99US-0145224. |
| PR | 27-JUL-1999; | 99US-0145276. |
| PR | 27-JUL-1999; | 99US-0145913. |
| PR | 28-JUL-1999; | 99US-0145919. |
| PR | 02-AUG-1999; | 99US-0145951. |
| PR | 02-AUG-1999; | 99US-0146386. |
| PR | 03-AUG-1999; | 99US-0146388. |
| PR | 04-AUG-1999; | 99US-0146389. |
| PR | 04-AUG-1999; | 99US-0147038. |
| PR | 05-AUG-1999; | 99US-0147204. |
| PR | 06-AUG-1999; | 99US-0147302. |
| PR | 06-AUG-1999; | 99US-0147192. |
| PR | 09-AUG-1999; | 99US-0147260. |
| PR | 10-AUG-1999; | 99US-0147303. |
| PR | 11-AUG-1999; | 99US-0147416. |
| PR | 12-AUG-1999; | 99US-0147493. |
| PR | 13-AUG-1999; | 99US-0147935. |
| PR | 13-AUG-1999; | 99US-0148171. |
| PR | 16-AUG-1999; | 99US-0148319. |
| PR | 17-AUG-1999; | 99US-0148341. |
| PR | 18-AUG-1999; | 99US-0148565. |
| PR | 20-AUG-1999; | 99US-0148684. |
| PR | 20-AUG-1999; | 99US-0149368. |
| PR | 23-AUG-1999; | 99US-0149175. |
| PR | 23-AUG-1999; | 99US-0149426. |
| PR | 25-AUG-1999; | 99US-0149722. |
| PR | 26-AUG-1999; | 99US-0149723. |
| PR | 27-AUG-1999; | 99US-0149929. |
| PR | 27-AUG-1999; | 99US-0149902. |
| PR | 30-AUG-1999; | 99US-0149930. |
| PR | 31-AUG-1999; | 99US-0150566. |
| PR | 01-SEP-1999; | 99US-0150884. |
| PR | 07-SEP-1999; | 99US-0151065. |
| PR | 10-SEP-1999; | 99US-0151066. |
| PR | 13-SEP-1999; | 99US-0151080. |
| PR | 15-SEP-1999; | 99US-0151303. |
| PR | 16-SEP-1999; | 99US-0151438. |
| PR | 20-SEP-1999; | 99US-0151930. |
| PR | 22-SEP-1999; | 99US-0152363. |
| PR | 23-SEP-1999; | 99US-0153070. |
| PR | 24-SEP-1999; | 99US-0153758. |
| PR | 28-SEP-1999; | 99US-0154018. |
| PR | 29-SEP-1999; | 99US-0154039. |
| PR | 04-OCT-1999; | 99US-0154779. |
| PR | 05-OCT-1999; | 99US-0155139. |
| PR | 06-OCT-1999; | 99US-0155486. |
| PR | 07-OCT-1999; | 99US-0155659. |
| PR | 08-OCT-1999; | 99US-0156458. |
| PR | 12-OCT-1999; | 99US-0156596. |
| PR | 13-OCT-1999; | 99US-0157117. |
| PR | 13-OCT-1999; | 0157753. |
| PR | 14-OCT-1999; | 0157865. |
| PR | 14-OCT-1999; | 0158029. |
| PR | 14-OCT-1999; | 0158232. |
| PR | | 0158369. |
| PR | | 0159293. |
| PR | | 0159294. |
| PR | | 0159295. |
| PR | | 0159329. |
| PR | | 0159330. |
| PR | | 0159331. |

```
PR   14-OCT-1999;  99US-0159637.
PR   14-OCT-1999;  99US-0159638.
PR   18-OCT-1999;  99US-0159584.
PR   21-OCT-1999;  99US-0160741.
PR   21-OCT-1999;  99US-0160767.
PR   21-OCT-1999;  99US-0160768.
PR   21-OCT-1999;  99US-0160770.
PR   21-OCT-1999;  99US-0160814.
PR   21-OCT-1999;  99US-0160815.
PR   22-OCT-1999;  99US-0160980.
PR   22-OCT-1999;  99US-0160981.
PR   22-OCT-1999;  99US-0160989.
PR   25-OCT-1999;  99US-0161404.
PR   25-OCT-1999;  99US-0161405.
PR   25-OCT-1999;  99US-0161406.
PR   26-OCT-1999;  99US-0161359.
PR   26-OCT-1999;  99US-0161360.
PR   26-OCT-1999;  99US-0161361.
PR   28-OCT-1999;  99US-0161920.
PR   28-OCT-1999;  99US-0161992.
PR   28-OCT-1999;  99US-0161993.
PR   29-OCT-1999;  99US-0162142.

ry Match               6.5%;  Score 30;  DB 21;  Length 1329;
Best Local Similarity  53.4%;  Pred. No. 4.5;
Matches  63;  Conservative  0;  Mismatches  55;  Indels  0;  Gaps  0;

Qy   165 gtatctcaatgaccaccatgccctccaaccatgccaactaatggcaatggccctgatcg 224
         || |||| ||   | || ||||| | ||||| | |||| | |  | ||||||
Db   296 GTTTCCGATCGACCACCTCGACTTCGATCCAACCATACCTCTCCCATACATGATCT 237

Qy   225 gatttcttaccattgtcggatttaccgcttatgggcttcatcaacagaggatgcc 282
         | ||||| || | ||| |||  |  | | ||||| || | ||  ||| |||
Db   236 AAAATCTCTCCGAAATTCCGAATCTCCCCGTCACCGGAGATCCACGCCGGTGAAAGCC 179

RESULT 12
AAF63962/c
ID   AAF63962 standard; DNA; 5616 BP.
XX
AC   AAF63962;
XX
DT   05-APR-2001  (first entry)
XX
DE   Human tankyrase1 3'UTR SEQ ID NO: 146.
XX
KW   Human; tankyrase2; TANK2; TRF1; telomere; cancer; neoplasm; aging;
KW   inflammatory disorder; ds.
XX
OS   Homo sapiens.
Y'
F.   WO200100849-A1.
XX
PD   04-JAN-2001.
XX
PF   28-JUN-2000; 2000WO-US17827.
XX
PR   29-JUN-1999;  99US-0141582.
XX
PA   (ICOS-) ICOS CORP.
XX
PI   Christenson E, Demaggio AJ, Goldman PS, McElligott DL;
XX
DR   WPI; 2001-102896/11.
XX
PT   New tankyrase2 polypeptides, useful for treating conditions mediated by
PT   poly(adenosine diphosphate-ribose) polymerase activity e.g. cancers,
PT   inflammatory and autoimmune disorders -
XX
PS   Example 5; Page 220-221; 242pp; English.
XX
CC   The present invention provides the protein and coding sequence for the
CC   human tankyrase2 protein. This is found in two different versions,
```

```
CC   designated TANK2-LONG and TANK2-SHORT. Tankyrase2 has
CC   polyADP-ribosylation activity and is involved in the modification of
CC   TRF1, which is a telomere-specific binding protein. The regulation of
CC   telomere length, in which TRF1 has a role, is linked to ageing and
CC   cancer. The sequences are useful in the treatment of cancers and
CC   inflammatory disorders.
XX
SQ   Sequence 5616 BP; 1612 A; 1076 C; 1059 G; 1869 T; 0 other;

Query Match             6.5%;  Score 30;  DB 22;  Length 5616;
Best Local Similarity  52.0%;  Pred. No. 8.9;
Matches  66;  Conservative  0;  Mismatches  61;  Indels  0;  Gaps  0;

Qy   290 caccctacccacacccgtgaccaacctggcctttaaataaacctatctttgtcatt 349
         ||| ||||| || || ||||| | ||| | |||| ||  | ||| ||| || |||||
Db   917 CACGCTCCTCACCCTCCTTAATGCAGTCTTCCTTCACAGCCTCCCAGTTTTTCACCT 858

Qy   350 cggataaactttcctgtcttcagcaaatgaagaaccccctgtcngtccttgggcggg 409
         ||| ||    || | | | || || | | ||| | | |  || | |||| ||| |||
Db   857 TGGAACTCAAGCCCGTGTCAACCTGCATCTCCAGTACTGTCCGGAAGACCTCTGTGGCAG 798

Qy   410 cctttg 416
         ||| ||
Db   797 GCTGTG 791

RESULT 13
AAC53576
ID   AAC53576 standard; DNA; 798 BP.
XX
AC   AAC53576;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 74984.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   23-APR-1999;  99US-0130891.
PR   28-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0133048.
PR   30-APR-1999;  99US-0132407.
PR   04-MAY-1999;  99US-0133484.
PR   05-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   06-MAY-1999;  99US-0132487.
PR   07-MAY-1999;  99US-0132863.
PR   11-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
```

us-09-521-640-2_1.rng

Wed Nov 7 10:16:52 2001

```
PR   14-MAY-1999;   99US-0134219.        PR   26-JUL-1999;   99US-0145276.
PR   14-MAY-1999;   99US-0134221.        PR   27-JUL-1999;   99US-0145913.
PR   14-MAY-1999;   99US-0134370.        PR   27-JUL-1999;   99US-0145918.
PR   18-MAY-1999;   99US-0134768.        PR   27-JUL-1999;   99US-0145919.
PR   19-MAY-1999;   99US-0134941.        PR   28-JUL-1999;   99US-0145951.
PR   20-MAY-1999;   99US-0135124.        PR   02-AUG-1999;   99US-0146386.
PR   21-MAY-1999;   99US-0135353.        PR   02-AUG-1999;   99US-0146388.
PR   24-MAY-1999;   99US-0135629.        PR   02-AUG-1999;   99US-0146389.
PR   25-MAY-1999;   99US-0136021.        PR   03-AUG-1999;   99US-0147038.
PR   27-MAY-1999;   99US-0136392.        PR   04-AUG-1999;   99US-0147204.
PR   28-MAY-1999;   99US-0136782.        PR   04-AUG-1999;   99US-0147302.
PR   01-JUN-1999;   99US-0137222.        PR   05-AUG-1999;   99US-0147192.
PR   03-JUN-1999;   99US-0137528.        PR   06-AUG-1999;   99US-0147260.
PR   04-JUN-1999;   99US-0137502.        PR   06-AUG-1999;   99US-0147303.
PR   07-JUN-1999;   99US-0137724.        PR   09-AUG-1999;   99US-0147416.
PR   0?-JUN-1999;   99US-0138094.        PR   09-AUG-1999;   99US-0147493.
PR   ?-JUN-1999;    99US-0138540.        PR   10-AUG-1999;   99US-0147935.
PR   ?-JUN-1999;    99US-0138847.        PR   11-AUG-1999;   99US-0148171.
PR   14-JUN-1999;   99US-0139119.        PR   12-AUG-1999;   99US-0148319.
PR   16-JUN-1999;   99US-0139452.        PR   13-AUG-1999;   99US-0148341.
PR   16-JUN-1999;   99US-0139453.        PR   13-AUG-1999;   99US-0148565.
PR   17-JUN-1999;   99US-0139492.        PR   16-AUG-1999;   99US-0148684.
PR   18-JUN-1999;   99US-0139454.        PR   17-AUG-1999;   99US-0149368.
PR   18-JUN-1999;   99US-0139455.        PR   18-AUG-1999;   99US-0149175.
PR   18-JUN-1999;   99US-0139456.        PR   20-AUG-1999;   99US-0149426.
PR   18-JUN-1999;   99US-0139457.        PR   20-AUG-1999;   99US-0149722.
PR   18-JUN-1999;   99US-0139458.        PR   20-AUG-1999;   99US-0149723.
PR   18-JUN-1999;   99US-0139459.        PR   23-AUG-1999;   99US-0149929.
PR   18-JUN-1999;   99US-0139460.        PR   25-AUG-1999;   99US-0149426.
PR   18-JUN-1999;   99US-0139461.        PR   26-AUG-1999;   99US-0149902.
PR   18-JUN-1999;   99US-0139462.        PR   27-AUG-1999;   99US-0149930.
PR   18-JUN-1999;   99US-0139463.        PR   27-AUG-1999;   99US-0150566.
PR   18-JUN-1999;   99US-0139750.        PR   30-AUG-1999;   99US-0150884.
PR   21-JUN-1999;   99US-0139763.        PR   31-AUG-1999;   99US-0151065.
PR   22-JUN-1999;   99US-0139817.        PR   01-SEP-1999;   99US-0151066.
PR   23-JUN-1999;   99US-0139899.        PR   07-SEP-1999;   99US-0151080.
PR   23-JUN-1999;   99US-0140353.        PR   10-SEP-1999;   99US-0151303.
PR   24-JUN-1999;   99US-0140354.        PR   13-SEP-1999;   99US-0151438.
PR   28-JUN-1999;   99US-0140695.        PR   15-SEP-1999;   99US-0151930.
PR   29-JUN-1999;   99US-0140823.        PR   16-SEP-1999;   99US-0152363.
PR   30-JUN-1999;   99US-0140991.        PR   20-SEP-1999;   99US-0153070.
PR   01-JUL-1999;   99US-0141287.        PR   22-SEP-1999;   99US-0153758.
PR   01-JUL-1999;   99US-0141842.        PR   23-SEP-1999;   99US-0154018.
PR   ?2-JUL-1999;   99US-0142154.        PR   24-SEP-1999;   99US-0154039.
PR   6-JUL-1999;    99US-0142055.        PR   28-SEP-1999;   99US-0154779.
PR   6-JUL-1999;    99US-0142390.        PR   29-SEP-1999;   99US-0155139.
PR   08-JUL-1999;   99US-0142803.        PR   04-OCT-1999;   99US-0155486.
PR   09-JUL-1999;   99US-0142920.        PR   05-OCT-1999;   99US-0155659.
PR   12-JUL-1999;   99US-0142977.        PR   06-OCT-1999;   99US-0156458.
PR   13-JUL-1999;   99US-0143542.        PR   07-OCT-1999;   99US-0156596.
PR   14-JUL-1999;   99US-0143624.        PR   08-OCT-1999;   99US-0157117.
PR   15-JUL-1999;   99US-0144005.        PR   12-OCT-1999;   99US-0157753.
PR   16-JUL-1999;   99US-0144085.        PR   13-OCT-1999;   99US-0157865.
PR   16-JUL-1999;   99US-0144086.        PR   13-OCT-1999;   99US-0158029.
PR   19-JUL-1999;   99US-0144325.        PR   14-OCT-1999;   99US-0158232.
PR   19-JUL-1999;   99US-0144331.        PR   14-OCT-1999;   99US-0158369.
PR   19-JUL-1999;   99US-0144332.        PR   14-OCT-1999;   99US-0159293.
PR   19-JUL-1999;   99US-0144333.        PR   14-OCT-1999;   99US-0159294.
PR   19-JUL-1999;   99US-0144334.        PR   14-OCT-1999;   99US-0159295.
PR   19-JUL-1999;   99US-0144335.        PR   18-OCT-1999;   99US-0159329.
PR   20-JUL-1999;   99US-0144352.        PR   21-OCT-1999;   99US-0159330.
PR   20-JUL-1999;   99US-0144632.        PR   21-OCT-1999;   99US-0159331.
PR   21-JUL-1999;   99US-0144884.        PR   21-OCT-1999;   99US-0159637.
PR   21-JUL-1999;   99US-0144814.        PR   21-OCT-1999;   99US-0159638.
PR   21-JUL-1999;   99US-0145086.        PR   21-OCT-1999;   99US-0159584.
PR   22-JUL-1999;   99US-0145088.        PR   22-OCT-1999;   99US-0160741.
PR   22-JUL-1999;   99US-0145085.        PR   22-OCT-1999;   99US-0160767.
PR   22-JUL-1999;   99US-0145087.        PR   22-OCT-1999;   99US-0160768.
PR   22-JUL-1999;   99US-0145089.        PR   21-OCT-1999;   99US-0160770.
PR   22-JUL-1999;   99US-0145192.        PR   21-OCT-1999;   ......160814.
PR   23-JUL-1999;   99US-0145145.        PR   21-OCT-1999;   99US-0160815.
PR   23-JUL-1999;   99US-0145218.        PR   22-OCT-1999;   ....-0160980.
PR   23-JUL-1999;   99US-0145224.        PR   22-OCT-1999;   ....0160981.
                                         PR   22-OCT-1999;   ..-0160989.
```

PR    25-OCT-1999;    99US-0161404.
PR    25-OCT-1999;    99US-0161405.
PR    25-OCT-1999;    99US-0161406.
PR    26-OCT-1999;    99US-0161359.
PR    26-OCT-1999;    99US-0161360.
PR    26-OCT-1999;    99US-0161361.
PR    28-OCT-1999;    99US-0161920.
PR    28-OCT-1999;    99US-0161992.
PR    28-OCT-1999;    99US-0161993.
PR    29-OCT-1999;    99US-0162142.
XX

Query Match             6.4%;    Score 29.6;  DB 21;   Length 798;
Best Local Similarity   50.0%;   Pred. No. 4.8;
Matches    74;  Conservative    0;  Mismatches    74;  Indels    0;  Gaps    0;

Qy      67  atccctttatgagcgcttaagtcactgccgcgcttacacgcggactgagaaacctgcgta  126
                |||||   ||    || || || |  | |||| ||| ||||| | || | |   || | |
Db     581  atccatgtcttaggtatgaagtagctgtgagcgcctttcttcgtcacttgatccagattg  640

Qy     127  ccgctaatcttcttcacacatccctatgcactggcggtatctaatgaccaccatcgc  186
                |    ||||| | ||||        || | |   ||||| |   |||  || ||| |
Db     641  ctgtttcctcatgtcacatttggatctagcctcggtttagttagtttcgcaaagaa  700

Qy     187  cttccaaccatgccaactaatggcaatg  214
                || ||||| | | ||  || | ||||
Db     701  ctcccaacttttctctcttcttggaatg  728

RESULT 14
AAC77435
ID    AAC77435 standard; cDNA; 1095 BP.
XX
AC    AAC77435;
XX
DT    08-FEB-2001  (first entry)
XX
DE    Human ORFX ORF2990 polynucleotide sequence SEQ ID NO:5979.
XX
KW    Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW    vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW    anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW    immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW    hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW    antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW    antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW    neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW    cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW    cholesterol ester storage; systemic lupus erythematosus; infection;
KW    severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW    allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW    bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW    thrombosis; contraceptive; ss.
XX
OS    Homo sapiens.
XX
PN    WO20005847-A2.
XX
PD    05-OCT-2000.
XX
PF    31-MAR-2000;  2000WO-US08621.
XX
PR    31-MAR-1999;  99US-0127607.
PR    02-APR-1999;  99US-0127636.
PR    05-APR-1999;  99US-0127728.
PR    30-MAR-2000;  2000US-0540763.
XX
PA    (CURA-) CURAGEN CORP.
XX
PI    Shimkets RA,  Leach M;
XX
DR    WPI; 2000-602362/57.
DR    P-PSDB; AAB43226.
XX

---

PT    Novel nucleic acids and peptides derived from open reading frame X,
PT    useful for treating e.g. cancers, proliferative disorders,
PT    neurodegenerative disorders and cardiovascular disease -
XX
PS    Claim 5; Page 5156-5157; 5507pp; English.
XX
CC    AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC    which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC    sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC    antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC    osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC    immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC    antidiabetic; hypotensive; dermatological; immunosuppressive;
CC    antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC    antithyroid; and antianaemic. The sequences can be used for determining
CC    the presence of or predisposition to, or preventing or treating
CC    pathological conditions associated with an ORFX-associated disorder. The
CC    nucleic acids can be used to express ORFX proteins in gene therapy
CC    vectors. The proteins and nucleic acids may be used to treat cancers,
CC    proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC    graft vs host disease, cardiovascular disease, diabetes mellitus,
CC    hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC    erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC    bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC    allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC    nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC    coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ    Sequence 1095 BP; 259 A; 333 C; 291 G; 210 T; 2 other;

Query Match             6.4%;    Score 29.6;  DB 21;   Length 1095;
Best Local Similarity   47.8%;   Pred. No. 5.6;
Matches    86;  Conservative    0;  Mismatches    94;  Indels    0;  Gaps    0;

Qy     179  accatcgcctccaaccatgccaactaatgcaatggccctgatgcgatctcttcttacca  238
                |||| ||  ||| || |||| | ||| ||  ||| ||| ||| | | ||     |||
Db     864  accacaaactcccctccaggctgccctgggacatggtgatgtgggtgcaggagcca  923

Qy     239  ttgtcggatttacccgctatggcgcttcatcaacagagatgcccagtatcacccctacc  298
               |||   | |||| |||   |||| |||| || || ||||||| |||  |||||  |||
Db     924  gtgtctgttgggactcgcaagtgcctcatcacagccaccccaccagtgtctc  983

Qy     299  cacaccccgtgaccaacttggcctttaaataaaacctatcttgtcatttcggaataaa  358
                || | |   ||||| |||  |||| ||    ||| | | ||   | ||  |||
Db     984  cccagtgcagactcaagttatgctgaaatgaaaagtctatctggtagtgggtaaaaa  1043

RESULT 15
AAC95117
ID    AAC95117 standard; cDNA; 640 BP.
XX
AC    AAC95117;
XX
DT    19-FEB-2001  (first entry)
XX
DE    Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:1612.
XX
KW    Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
KW    flea infestation; vaccine; antiparasitic; therapeutic target;
KW    diagnosis; detection; ss.
XX
OS    Ctenocephalides felis.
XX
PN    WO20061621-A2.
XX
PD    19-OCT-2000.
XX
PF    07-APR-2000;  2000WO-US09437.
XX
PR    09-APR-1999;  99US-0128704.
XX
PA    (HESK-) HESKA CORP.

XX
PI    Brandt KS, Gaines PJ, Stinchcomb DT, Wisnewski N;
XX
DR    WPI; 2000-656323/63.
XX
PT    Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT    acids useful for the prevention, diagnosis and treatment of flea
PT    infestations -
XX
PS    Claim 26; Page 784; 964pp; English.
XX
CC    The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC    acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC    or head and nerve cord (HNC) tissue. The invention also relates to the
CC    encoded proteins. The invention additionally encompasses expression
CC    constructs, recombinant viruses and recombinant cells comprising the
CC    nucleic acids of the invention, recombinant production of the proteins,
CC    antibodies against the proteins, a method of identifying inhibitors of
CL    the proteins, and compositions comprising the inhibitors for
CC    administration to an animal. The nucleic acids, and the proteins they
CC    encode may be used in the prevention, treatment and diagnosis of diseases
CC    associated with flea infestations. For example, the nucleic acids may be
CC    used to produce an HMT or HNC protein according to standard recombinant
CC    DNA methodology by inserting the nucleic acids into a host cell and
CC    culturing the cell to express the protein. The HMT and HNC nucleic acids
CC    may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC    and quantitate the presence of cat flea or other homologous nucleic acid
CC    sequences in samples. They may also be used to study the expression and
CC    function of the proteins and their role in metabolism. The HMT and HNC
CC    proteins may be used as antigens in the production of specific
CC    antibodies, and in assays to identify modulators (agonists and
CC    antagonists) of HMT and/or HNC protein expression and activity. The
CC    anti-HMT/HNC protein antibodies and antagonists may also be used to
CC    downregulate protein expression and activity. The antibodies may also be
CC    used as diagnostic agents for detecting the presence of flea polypeptides
CC    in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC    present sequence represents a cat flea HMT cDNA of the invention.
XX
SQ    Sequence 640 BP; 208 A; 135 C; 133 G; 153 T; 11 other;

Query Match             6.3%;   Score 29.4;  DB 21;   Length 640;
Best Local Similarity   53.3%;          Pred. No. 5.1;
Matches    96;   Conservative    0;   Mismatches    80;   Indels    4;   Gaps    2;

Qy      37 agactcgctgtaagacctcggtccagccgaatcccctttatagagcgcttaagtcactgccg   96
           ||  ||||| |||  ||||| ||| || |  || |||||   || ||| ||||| ||||| ||
Db     452 aaactcgaggggggggcgcgtcccaattcgcctatagtgagtcgtatacaattcactggcg  511

Qy      97 cgcttacacg--cggactgagaaacctgcgtaccgctaatcctccacacatccccctat  154
           || ||||||| ||||||| |||||  |||||||||| |||| || ||| || |||| |
Db     512 cgtttacacgtcgtgctgggaaaccctgcgtaccacttatcgcttgnacacatcccctt  571

Qy     155 cgcactggcggtatctcaatgaccaccatcgcttccaacatgcaactaatggcaatg  214
           | || ||  ||||  |||| || |||| ||||| |||| || |||||| ||| |||
Db     572 cncantg--cgtatangaaagcccncgatcgncttnacagtgccanctgatgggaang  629

Search completed: November 7, 2001, 04:11:03
Job time: 2723 sec

OM nucleic - nucleic search, using sw model

```
                 GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.
```

Run on:          November  7, 2001, 03:23:35 ; Search time 1170.27 Seconds
                                              (without alignments)
                                              3756.037 Million cell updates/sec

Title:           US-09-521-640-2
Perfect score:   465
Sequence:        1 ggggagtatcgtgagttcac........ccctctgttcaaaacacn 465

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Sec..ned:        10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters:        20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                   1:  gb_est1:*
                   2:  gb_est2:*
                   3:  gb_est3:*
                   4:  gb_est4:*
                   5:  gb_est5:*
                   6:  gb_est6:*
                   7:  gb_est7:*
                   8:  gb_est8:*
                   9:  gb_est9:*
                  10:  gb_est10:*
                  11:  gb_est11:*
                  12:  gb_est12:*
                  13:  gb_est13:*
                  14:  gb_est14:*
                  15:  gb_est15:*
                  16:  gb_est16:*
                  17:  gb_est17:*
                  18:  gb_est18:*
                  19:  gb_est19:*
                  20:  gb_est20:*
                  21:  gb_est21:*
                  22:  gb_est22:*
                  23:  gb_est23:*
                  24:  gb_est24:*
                  25:  gb_est33:*
                  26:  gb_est34:*
                  27:  gb_est35:*
                  28:  gb_est36:*
                  29:  gb_est37:*
                  30:  gb_est38:*
                  31:  gb_est39:*
                  32:  gb_est40:*
                  33:  em_estba:*
                  34:  em_estfun:*
                  35:  em_esthum1:*
                  36:  em_esthum2:*
                  37:  em_esthum3:*
                  38:  em_esthum4:*
                  39:  em_esthum5:*
                  40:  em_esthum6:*
                  41:  em_esthum7:*
                  42:  em_esthum8:*
                  43:  em_esthum9:*
                  44:  em_esthum10:*
                  45:  em_esthum11:*
                  46:  em_esthum12:*
                  47:  em_esthum13:*
                  48:  em_esthum14:*
                  49:  em_esthum15:*
                  50:  em_esthum16:*
                  51:  em_esthum17:*
                  52:  em_esthum18:*
                  53:  em_esthum19:*
                  54:  em_esthum20:*
                  55:  em_esthum21:*
                  56:  em_esthum22:*
                  57:  em_esthum23:*
                  58:  em_esthum24:*
                  59:  em_esthum25:*
                  60:  em_esthum26:*
                  61:  em_esthum27:*
                  62:  em_esthum28:*
                  63:  em_estin1:*
                  64:  em_estin2:*
                  65:  em_estin3:*
                  66:  em_estin4:*
                  67:  em_estin5:*
                  68:  em_estom1:*
                  69:  em_estom2:*
                  70:  em_estov1:*
                  71:  em_estov2:*
                  72:  em_estpl1:*
                  73:  em_estpl2:*
                  74:  em_estpl3:*
                  75:  em_estpl4:*
                  76:  em_estpl5:*
                  77:  em_estpl6:*
                  78:  em_estpl7:*
                  79:  em_estpl8:*
                  80:  em_estpl9:*
                  81:  em_estpl10:*
                  82:  em_estro1:*
                  83:  em_estro2:*
                  84:  em_estro3:*
                  85:  em_estro4:*
                  86:  em_estro5:*
                  87:  em_estro6:*
                  88:  em_estro7:*
                  89:  em_estro8:*
                  90:  em_estro9:*
                  91:  em_estro10:*
                  92:  em_estro11:*
                  93:  em_estro12:*
                  94:  em_estro13:*
                  95:  em_estro14:*
                  96:  em_estro15:*
                  97:  em_estro16:*
                  98:  em_estro17:*
                  99:  em_estro18:*
                 100:  em_estro19:*
                 101:  em_estro20:*
                 102:  gb_est25:*
                 103:  gb_est26:*
                 104:  gb_est27:*
                 10 :  gb_est28:*
                 10 :  _est29:*
                 10 :  _est30:*
                 10 :  _est31:*
                 10 :  _est32:*
                 11 :  _est41:*
                 11 :  _est42:*
                 11 :  _est43:*
                 11 :  _est44:*
                 11 :  _est45:*
                 11 :  _est46:*
                 11 :  _est47:*
```

117:  gb_est48:*
118:  gb_est49:*
119:  gb_est50:*
120:  gb_est51:*
121:  gb_est52:*
122:  gb_est53:*
123:  gb_est54:*
124:  gb_est55:*
125:  gb_est56:*
126:  gb_est57:*
127:  gb_est58:*
128:  gb_est59:*
129:  gb_est60:*
130:  gb_est61:*
131:  gb_est62:*
132:  gb_est63:*
133:  gb_est64:*
134:  gb_est65:*
135:  gb_est66:*
136:  gb_est75:*
137:  gb_est76:*
138:  gb_est77:*
139:  gb_est78:*
140:  gb_est79:*
141:  gb_est80:*
142:  gb_est81:*
143:  gb_est82:*
144:  gb_est83:*
145:  gb_est84:*
146:  gb_est85:*
147:  gb_est86:*
148:  gb_est87:*
149:  gb_est88:*
150:  gb_est89:*
151:  gb_est90:*
152:  gb_est99:*
153:  gb_est100:*
154:  gb_est101:*
155:  gb_est102:*
156:  gb_est103:*
157:  gb_est104:*
158:  gb_est105:*
159:  gb_est106:*
160:  gb_est67:*
161:  gb_est68:*
162:  gb_est69:*
163:  gb_est70:*
164:  gb_est71:*
165:  gb_est72:*
166:  gb_est73:*
167:  gb_est74:*
168:  gb_est91:*
169:  gb_est92:*
170:  gb_est93:*
171:  gb_est94:*
172:  gb_est95:*
173:  gb_est96:*
174:  gb_est97:*
175:  gb_est98:*
176:  em_esthum29:*
177:  em_esthum30:*
178:  em_esthum31:*
179:  em_esthum32:*
180:  em_esthum33:*
181:  em_estom3:*
182:  em_estpl11:*
183:  em_estro21:*
184:  em_estro22:*
185:  em_estro23:*
186:  em_htc:*
187:  gb_est107:*
188:  gb_est108:*
189:  gb_est109:*

190:  gb_est110:*
191:  gb_est111:*
192:  gb_htc:*
193:  em_gss_fun:*
194:  em_gss_hum1:*
195:  em_gss_hum2:*
196:  em_gss_hum3:*
197:  em_gss_hum4:*
198:  em_gss_hum5:*
199:  em_gss_hum6:*
200:  em_gss_hum7:*
201:  em_gss_hum8:*
202:  em_gss_hum9:*
203:  em_gss_inv1:*
204:  em_gss_inv2:*
205:  em_gss_inv3:*
206:  em_gss_other:*
207:  em_gss_pln1:*
208:  em_gss_pln2:*
209:  em_gss_pro:*
210:  em_gss_rod1:*
211:  em_gss_rod2:*
212:  em_gss_rod3:*
213:  em_gss_rod4:*
214:  em_gss_rod5:*
215:  em_gss_vrt1:*
216:  em_gss_vrt2:*
217:  em_gss_vrt3:*
218:  gb_gss1:*
219:  gb_gss2:*
220:  gb_gss3:*
221:  gb_gss4:*
222:  gb_gss5:*
223:  gb_gss6:*
224:  gb_gss7:*
225:  gb_gss8:*
226:  gb_gss9:*
227:  gb_gss10:*
228:  gb_gss11:*
229:  gb_gss12:*
230:  gb_gss13:*
231:  gb_gss14:*
232:  gb_gss15:*
233:  gb_gss16:*
234:  gb_gss17:*
235:  gb_gss18:*
236:  gb_gss19:*
237:  gb_gss20:*
238:  gb_gss21:*
239:  gb_gss22:*
240:  gb_gss23:*
241:  gb_gss24:*
242:  gb_gss25:*
243:  gb_gss26:*
244:  gb_gss27:*
245:  gb_gss28:*
246:  gb_gss29:*
247:  gb_gss30:*
248:  gb_gss31:*
249:  gb_gss32:*
250:  gb_gss33:*
251:  gb_gss34:*
252:  em_gss_inv4:*
253:  em_gss_rod6:*
254:  em_gss_rod7:*
255:  em_gss_rod8:*
256:  gb_gss35:*
257:  gb_gss36:*
258:  gb_gss37:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 89.2 | 19.2 | 579 | 225 | AQ210729 | AQ210729 HS_2229_A |
| 2 | 64 | 13.8 | 489 | 233 | AQ761577 | AQ761577 HS_3187_A |
| 3 | 63.8 | 13.7 | 468 | 233 | AQ806004 | AQ806004 HS_3025_A |
| 4 | 60.6 | 13.0 | 740 | 235 | AQ914869 | AQ914869 nbeb0050K |
| 5 | 53.6 | 11.5 | 397 | 225 | AQ208121 | AQ208121 HS_3220_A |
| 6 | 50.2 | 10.8 | 1195 | 162 | BE035400 | BE035400 M004D10 M |
| 7 | 45.6 | 9.8 | 677 | 32 | AV702190 | AV702190 AV702190 |
| 8 | 45.2 | 9.7 | 173 | 226 | AQ304554 | AQ304554 HS_3247_B |
| 9 | 44.4 | 9.5 | 300 | 167 | BE401707 | BE401707 CNW02ELO2 |
| 10 | 43.6 | 9.4 | 205 | 233 | AQ763115 | AQ763115 HS_316l_A |
| 11 | 43.2 | 9.3 | 736 | 32 | AV721604 | AV721604 AV721604 |
| | 43 | 9.2 | 946 | 32 | AV726956 | AV726956 AV726956 |
| 13 | 42.8 | 9.2 | 168 | 234 | AQ823152 | AQ823152 HS_3186_B |
| 14 | 42.6 | 9.2 | 659 | 32 | AV705287 | AV705287 AV705287 |
| 15 | 42.4 | 9.1 | 312 | 256 | B45121 | B45121 HS-1060-B1- |
| 16 | 42.2 | 9.1 | 259 | 233 | AQ775293 | AQ775293 HS_315_A |
| 17 | 42.2 | 9.1 | 1101 | 219 | CNS00LT2 | CNS00LT2 Drosophil |
| 18 | 41.4 | 8.9 | 419 | 224 | AQ116709 | AQ116709 HS_3080_A |
| 19 | 41.4 | 8.9 | 818 | 32 | AV726386 | AV726386 AV726386 |
| 20 | 41 | 8.8 | 545 | 139 | BE755460 | BE755460 209303 MA |
| 21 | 40.6 | 8.7 | 982 | 15 | AI068596 | AI068596 mgae0003b |
| 22 | 40.4 | 8.7 | 403 | 233 | AQ785487 | AQ785487 HS_3063_A |
| 23 | 40.4 | 8.7 | 714 | 104 | AI965418 | AI965418 sc71e07.y |
| 24 | 40.4 | 8.7 | 847 | 32 | AV726234 | AV726234 AV726234 |
| 25 | 40.2 | 8.6 | 981 | 15 | AI068440 | AI068440 mgae0002b |
| 26 | 40 | 8.6 | 146 | 235 | AQ891463 | AQ891463 HS_3111_B |
| 27 | 40 | 8.6 | 432 | 224 | AQ139216 | AQ139216 HS_3090_A |
| 28 | 40 | 8.6 | 961 | 112 | BG343702 | BG343702 HVSMEg000 |
| 29 | 39.6 | 8.5 | 824 | 112 | AW155132 | AW155132 mqie00020 |
| 30 | 39.6 | 8.5 | 870 | 32 | AV726543 | AV726543 AV726543 |
| 31 | 39.4 | 8.5 | 920 | 32 | AV727278 | AV727278 AV727278 |
| 32 | 39.2 | 8.4 | 819 | 153 | BG418812 | BG418812 HVSMEk002 |
| 33 | 39 | 8.4 | 768 | 15 | AI068832 | AI068832 mgae0004b |
| 34 | 38.8 | 8.3 | 457 | 233 | AQ775019 | AQ775019 HS_3155_A |
| 35 | 38.8 | 8.3 | 890 | 164 | BE214026 | BE214026 HV_CEb000 |
| 36 | 38.8 | 8.3 | 985 | 162 | BE034549 | BE034549 MK01B09 M |
| 37 | 38.6 | 8.3 | 1200 | 110 | AV761384 | AV761384 AV761384 |
| 38 | 38.2 | 8.2 | 929 | 152 | BG343386 | BG343386 HVSMEg000 |
| 39 | 38 | 8.2 | 972 | 106 | AL576318 | AL576318 AL576318 |
| | 37.8 | 8.1 | 287 | 233 | AQ805901 | AQ805901 HS_3235_A |
| 1 | 37.8 | 8.1 | 1017 | 162 | BE039690 | BE039690 OC02E04 O |
| 42 | 37.6 | 8.1 | 430 | 226 | AQ301842 | AQ301842 HS_3195_A |
| 43 | 37.6 | 8.1 | 945 | 164 | BE195997 | BE195997 HVSMEh009 |
| 44 | 36.8 | 7.9 | 263 | 120 | AW800405 | AW800405 MR2-UM006 |
| 45 | 36.8 | 7.9 | 802 | 32 | AV725250 | AV725250 AV725250 |

## ALIGNMENTS

RESULT 1
AQ210729
LOCUS        AQ210729        579 bp    DNA        GSS        18-SEP-1998
DEFINITION   HS_2229_Al_C03_MR CIT Approved Human Genomic Sperm Library D Homo
             sapiens genomic clone Plate-2229 Col=5 Row=E, DNA sequence.
ACCESSION    AQ210729
VERSION      AQ210729.1    GI:3619698
KEYWORDS     GSS.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 579)
AUTHORS      Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
             Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
             Hood,L.
TITLE        Sequence-tagged connectors: A sequence approach to mapping and

JOURNAL
MEDLINE
COMMENT      Contact: Mahairas GG, Wallace JC, Hood L
             High Throughput Sequencing Center
             University of Washington
             401 Queen Anne Avenue North, Seattle, WA 98109, USA
             Tel: (206) 616-3618
             Fax: (206) 616-3887
             Email: jwallace@u.washington.edu
             Sequence Tagged Connector
             Plate: 2229  row: E  column: 5
             Class: BAC ends
             High quality sequence stop: 579.
FEATURES             Location/Qualifiers
     source          1..579
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2229 Col=5 Row=E"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT    148 a    181 c    111 g    138 t    1 others
ORIGIN

Query Match           19.2%;    Score 89.2;  DB 225;  Length 579;
Best Local Similarity 68.5%;    Pred. No. 4e-17;
Matches 152;  Conservative 0;  Mismatches 68;  Indels 2;  Gaps 2;

Qy      92 tgccgcgcttacacgcggactgagaaacctgctaccgctaatcttcacacatcccc 151
           ||||| ||||||||| | |||||| || ||||||| ||| ||||||||||||||||||
Db      89 TGCCTGCGTTACACGCGACGGAAAACCCTGCGTTCCCACTTATCGCTGCCACACATCCCC 148

Qy     152 tatccactggcggtatctcaatgaccaccatcg-ccttccaacatgccaactaatggc 210
           | |||||||| ||| || | || | |||||| ||| ||||||| ||||| |||| |||||
Db     149 TTTCGCACTGCGTAATACGAAAGACCCGCCCATCGGCCCTTCCAACAGTGCCACTGAATGGC 208

Qy    ·211 aatggccctgatcggattctcttaccattgtcggatttac-ccgctatggcgcttcatc 269
           |||||||| ||||| |||||| | ||| || | ||| |||| || |||| |||| |||
Db     209 AATGCGCCCATGCCGATTTCTCTTACCTCTGGCGGATTTTACACCGATAGGGCCTCTCATC 268

Qy     270 aacagaggatgccctagtatcaccctaccacacccgtgac 311
           ||| | |||||||||| || ||||| || | ||||| |||
Db     269 ATCTGCTCTGATGCCGATAGTAACCGGCCGACCCCGCAACAC 310

RESULT 2
AQ761577
LOCUS        AQ761577        489 bp    DNA        GSS        27-JUL-1999
DEFINITION   HS_3187_Al_G06_MR CIT Approved Human Genomic Sperm Library D Homo
             sapiens genomic clone Plate-3187 Col-11 Row-M, DNA sequence.
ACCESSION    AQ761577
VERSION      AQ761577.1    GI:5627380
KEYWORDS     GSS.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 489)
AUTHORS      Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
             Keller,A. Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
             Hood,L.
TITLE        Sequence-tagged connectors: A sequence approach to mapping and
             scanning the human genome
JOURNAL
MEDLINE      Proc. N     Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
             993805P9
COMMENT      Contact:  ahairas GG, Wallace JC, Hood L
             High Th    hput Sequencing Center
             Univers    f Washington
             401 Que   nne Avenue North, Seattle, WA 98109, USA
             Tel: (2   6-3618

Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3187  row: M  column: 11
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 489.

FEATURES
     source          Location/Qualifiers
                     1..489
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3187 Col=11 Row=M"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                        E-coli DH10B"

BASE COUNT        98 a       153 c       104 g       128 t       6 others
ORIGIN

Query Match            13.8%;      Score 64;    DB 233;     Length 489;
t Local Similarity     62.0%;      Pred. No. 3.1e-09;
ches 168:  Conservative  0;   Mismatches 96;   Indels  7;   Gaps  4;

Qy   83  ttagtcactgacggcggcttacacgcggact-gagaaacctgctaccgctaatctctc 141
             ||||||| |||||||| |||||||||||||| || | ||||| || |||||||| |
Db  123  TCAATTACTGGCCGCGCCTACACGCGAACTGGAAAACCCTGCGTTCCACTAATCGCTG 182

Qy  142  acacatccctatcgcactggcggtatccatgaccacatcgcttccaaccatgcca 201
             |||||||||| |  |||||| |||| |||| ||||| || |||| ||||
Db  183  ACACATCCCTTTCC---CACTGGTATACGAAAGGCCCCGACCCCTTCCAGTGCCA 239

Qy  202  actaatgacatggcccctgatcggatttcttctaccattgtcggattaccgctatggc 261
             ||||| ||| ||||||||  ||| ||||||| ||||||| ||| || ||| ||||||
Db  240  CTGAT--GCAATGGCCCTG-TCGGATTTCTTCTACCATTCGGCGGATTCAACCCATTGGTGC 296

Qy  262  gcttcatcaacagaggatgccctagtatcacocctacccacccgtgaccaaccttggc 321
             |||||||| ||||||| ||||  ||| ||| |||||| ||||| ||| |||||||
Db  297  CTCCAACAACTGCTATCCGCTTTTAACACCGCACCGCACACCGTGACCAACNCTGGT 356

Qy  322  cttaaataaacctatctttgtcattcgg 352
             |||| ||| |||| ||| || || ||
Db  357  GTCTTCATGACACATTTTATTTGGTACATGGG 387

RESULT  3
AQ806004
LOCUS       AQ806004      468 bp    DNA                    GSS      09-AUG-1999
DEFINITION  HS_3025_A2_H11_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3025 Col=22 Row=O, DNA sequence.
ACCESSION   AQ806004
VERSION     AQ806004.1   GI:5723266
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 468)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu

---

Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3025  row: O  column: 22
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 468.

FEATURES
     source          Location/Qualifiers
                     1..468
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3025 Col-22 Row=O"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                        E-coli DH10B"

BASE COUNT       105 a       155 c        84 g       120 t        4 others
ORIGIN

Query Match            13.7%;      Score 63.8;   DB 233;   Length 468;
Best Local Similarity  58.7%;      Pred. No. 3.5e-09;
Matches 166:  Conservative  0;   Mismatches 112;   Indels  5;   Gaps  3;

Qy   12  tgagttcactcaatattagccctcagactcgctgtaagacctcgtccagccgaatccc 71
             ||||||| ||  ||||||||  || | |||||| || ||| ||||||||
Db  166  TGAGTTGCCACTTGATAGCCCTTCACCCCTCATATGTGCCAGTTGGAATTCCCCTATACG 225

Qy   72  tttatgagcgcttaagtcactgcgcgcgcttacacgcggactgagaaacctgcgtacc--- 128
             ||||||||| ||||| || |||| ||||||| || |||||||
Db  226  AGTCGATTACATTCACTGGCCGTCTTTACAACGTCGGACTCGAAAACCCTCGCGTTCCAA 285

Qy  129  cgctaatctcttcacacatcccctatcgcactgggtgtatctcaatg-accacatcgcc 187
             |||||||||  |  | ||||||||||  ||||||| || | | |||||||||||||
Db  286  CTTAATCCCTTGCACACATCCCCTTCGCCACTGGCTGTATACTAAAGGCCCCACCATCGCT 345

Qy  188  ttccaacatgccaactaatggcaatggccctgatcggattctcttaccattgtcggat 247
             || | |||||   ||||||| ||||||||||||| |||||||||| ||||||||| |
Db  346  CTCCCACAATGCCACTGAATGGCAA-GGCCCTGAGCCGCGATTTCTCTTACCATTGTCGGTT 404

Qy  248  ttaccgctatggcgcttcatcaacagagagatgccctagtatc 290
             || ||| ||||| ||||| ||| | ||||| ||| ||| |||
Db  405  TCACACGATTGGGCACCTAGACCAACTGTCTAGCCGATAGTAGC 447

RESULT  4
AQ914869
LOCUS       AQ914869     740 bp    DNA                     GSS      02-DEC-1999
DEFINITION  nbeb0050K20r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone nbeb0050K20r, DNA sequence.
ACCESSION   AQ914869
VERSION     AQ914869.1   GI:6511385
KEYWORDS    GSS.
SOURCE      Oryza sativa
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 740)
AUTHORS     Wing,R.A. and Dean,R.A.
TITLE       A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL     Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GGAAACAGCTATGACCATG
            Class: BAC ends
            High quality sequence start: 41
            High quality sequence stop: 190.

FEATURES
                     Location/Qualifiers

```
source       1..740
             /organism="Oryza sativa"
             /strain="Japonica"
             /cultivar="Nipponbare"
             /db_xref="taxon:4530"
             /clone="nbeb0050K20r"
             /clone_lib="CUGI Rice BAC Library (EcoRI)"
             /tissue_type="Leaf"
             /lab_host="E. coli DH10B"
             /note="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
```

Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximatley 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

```
BASE COUNT   195 a   173 c   190 g   181 t    1 others
ORIGIN

Query Match             13.0%;     Score 60.6;  DB 235;  Length 740;
Best Local Similarity   61.7%;     Pred. No. 4e-08;
Matches  113;  Conservative   0;  Mismatches  69;  Indels  1;  Gaps  1;

Qy    139 ttcacacaatcccctatcgcactggcgtatctcaatgaccacatcgccttccaccatg 198
          | || |||||||||| ||||   | ||| |||||||  |    ||  ||
Db    342 TGCAGACACATCCCCTTCGCAGCTGGCGTATAGCAAAGGCCGACCGATCGCCTTCACAGTG 401

Qy    199 ccaactaatggcccctgatcggatttctcttccattgtcggatttaccccgctat 258
          | ||| |        ||| |  | |||| || || ||  ||  | ||
Db    402 CGCACTGAATGGGAAGGGCCTGAGCGATTTTCTTTACCCATTGGCGGATTTACACGATAT 461

Qy    259 gggcttcatcaacagagaggatgccctagtatca-ccctacccaccacccgtgaccacct 317
          |||  ||  ||   |  | |||  |  ||| | ||| ||   | |||||||||||
Db    462 GGCCTCTAGACAATTGTCTGAGCGATAGTAGCAGCCCGAACCGCACACCGTGACCAACCC 521

Qy    318 tgg 320
          | |
Db    522 TTG 524
```

RESULT 5
AQ208121
```
LOCUS       AQ208121    397 bp   DNA             GSS   18-SEP-1998
DEFINITION  HS_3220_A2_H07_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate-3220 Col-14 Row-O, DNA sequence.
ACCESSION   AQ208121
VERSION     AQ208121.1  GI:3620856
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 397)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3220 row: O column: 14
            Class: BAC ends
            High quality sequence stop: 397.
FEATURES             Location/Qualifiers
     source          1..397
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate-3220 Col-14 Row-O"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT    93 a   111 c    85 g   106 t    2 others
ORIGIN

Query Match             11.5%;     Score 53.6;  DB 225;  Length 397;
Best Local Similarity   81.6%;     Pred. No. 5.4e-06;
Matches   62;  Conservative   0;  Mismatches 14;  Indels   0;  Gaps   0;

Qy    178 caccatcgccttccaacatgccaactaatggcaatggccctgatcggatttctcttacc 237
          | |||| |||||| || | |||| || |||| | |||||||||||| ||||||||
Db    162 CCCGATCCCCTTCCAACAGTGCAACAGTGCCAGCTGATCGATGGCAATGGCCCTGATGCGGCGATTTCTTTACC 221

Qy    238 attgtcggatttaccc 253
          |||||  | || | |||
Db    222 ATTGTGCGGATTACAC 237
```

RESULT 6
BE035400
```
LOCUS       BE035400    1195 bp    mRNA                 EST   07-JUN-2000
DEFINITION  M004D10 MO Mesembryanthemum crystallinum cDNA 5' similar to
            ribosomal protein 117, mRNA sequence.
ACCESSION   BE035400
VERSION     BE035400.1  GI:8330409
KEYWORDS    EST.
SOURCE      common ice plant.
ORGANISM    Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 1195)
AUTHORS     Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
            ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
            Scara,G., Wheeler,M. and Zepeda,G.R.
TITLE       Functional Genomics of Plant Stress Tolerance
JOURNAL     Unpublished (2000)
COMMENT     Contact: Michalowski,C.B.
            University of Arizona
            Bio Sciences West room 513, Tucson, AZ 85721, USA
            Tel: 520-621-7982
            Fax: 5?? 21-1697
            Email: ...@...arizona.edu
            An open ...ding frame exists.
FEATURES             Location/Qualifiers
     source          1..1195
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
                     /clone_lib="MO"
                     /tissue_type="apical meristem and leaf primordia"
                     /dev_stage="5 weeks"
                     /note="no stress"
```

RESULT 8
AQ304554
```
LOCUS       AQ304554    173 bp    DNA                    GSS        16-DEC-1998
DEFINITION  HS_3247_B2_A10_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3247 Col=20 Row=B, DNA sequence.
ACCESSION   AQ304554
VERSION     AQ304554.1  GI:4024340
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3247 row: B  column: 20
            Class: BAC ends
            High quality sequence stop: 173.
FEATURES             Location/Qualifiers
source          1..173
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Plate=3247 Col=20 Row=B"
                /clone_lib="CIT Approved Human Genomic Sperm Library D"
                /sex="male"
                /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                E-Coli DH10B"
BASE COUNT    53 a      53 c      32 g      35 t
ORIGIN
```

Query Match            9.7%;       Score 45.2;  DB 226;  Length 173;
Best Local Similarity  66.7%;      Pred. No. 0.0019;
Matches   80;  Conservative    0;  Mismatches   38;  Indels   2;  Gaps   1;

Qy    152 tatcgcactggcggtatctcaa--tgaccaccatgcgcttccaccatgccaactaatgg 209
          |||||||| ||||| || |||   || ||| ||| ||||| |||| ||| ||| ||||||
Db     53 TATCGCAGTGGCGATAACCAAAATGCACGACCGATCCCCTTCAACATTTCGCACTGATGG 112

Qy    210 caatggccctgatcggagtttctcttaccattgtcggattaccgctatggcggcttcatc 269
          || || |||| ||| || ||| ||||  || || || |||| |||| |||| | |||| ||
Db    113 CAACGGGCCTGAGCGGGATTTTCCTCACCACCACGAGCGGATCTACCCCATATAGAGCTTACTC 172

RESULT 9
BE401707
```
LOCUS       BE401707    300 bp    mRNA                   EST        21-JUL-2000
DEFINITION  CNW02EL021 ITEC CNW Wheat Powdery Mildew Resistant Library Triticum
            aestivum cDNA clone CNW02EL021, mRNA sequence.
ACCESSION   BE401707
VERSION     BE401707.1  GI:9361175
KEYWORDS    EST.
SOURCE      bread wheat.
ORGANISM    Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 300)
```

BASE COUNT   305 a     266 c     301 g     304 t     19 others
ORIGIN

Query Match           10.8%;       Score 50.2;  DB 162;  Length 1195;
Best Local Similarity 66.4%;       Pred. No. 8.3e-05;
Matches   87;  Conservative    0;  Mismatches   43;  Indels   1;  Gaps   1;

Qy     88 tcactgcgcgcttacacg-cggactgagaaacctggtaccgctaatctcttcacaca 146
          | |||||||  |||||||  |||||  ||||||||||||||||| |||| | ||| |||
Db    874 TAACTGCGCGTCGTTACACGTCTGACTGGGAAACCTGCGTACCCACTAAATGCTTGAGACA 933

Qy    147 tccccatcgcactggcggtatccaatgaccaccatcgcccttccaaccatgccaactaa 206
          |||| || |||||| | ||||| ||  || |||||| ||| || || ||||| |
Db    934 TCCCCTTTTGCGGTGGCGAATTCGAAAAGCCCGACGACGCGTTTTCACAGTTGCAACTGAT 993

Qy    207 tgcaatggcc 217
          || ||| |||
Db    994 TGCTATGGACC 1004

RESULT 7
AV702190
```
LOCUS       AV702190    677 bp    mRNA                   EST        08-OCT-2000
DEFINITION  AV702190 ADB Homo sapiens cDNA clone ADBCOH01 5', mRNA sequence.
ACCESSION   AV702190
VERSION     AV702190.1  GI:10718520
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 677)
AUTHORS     Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao
            ,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z.,
            Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
            ,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
TITLE       Homo sapiens cDNA ADB clones
JOURNAL     Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            20120, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
source          1..677
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="ADBCOH01"
                /clone_lib="ADB"
                /tissue_type="Adrenal gland"
                /dev_stage="Adult"
                /lab_host="SOLR"
                /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                XhoI"
BASE COUNT   166 a     161 c     156 g     194 t
ORIGIN
```

Query Match            9.8%;       Score 45.6;  DB 32;  Length 677;
Best Local Similarity 66.7%;       Pred. No. 0.002;
Matches   80;  Conservative    0;  Mismatches   39;  Indels   1;  Gaps   1;

Qy     93 gccgcgcttacacg-cggactgagaaacctggtaccgctaatctcttcacacatcccc 151
          ||||||||||||||  |||||  ||||||||||||||||| |||| | ||| |  |||||
Db    550 GCCGCGGTTTACACGCCTGACTGGGAAACCTTGGCGTACCCACTAATCGCTTGAGACAATCCC 609

Qy    152 tatcgcactggcggtatctcaatgaccaccatcgcccttccaaccatgccaactaatgca 211
          | ||||||||||||| || |||||| ||| ||| |  ||| || |||| |   ||||| ||
Db    610 TTTGCAGTTGCGAATAGCCAAAGGCCCGACCGATGGCTTTCCACAGTTGCGCAGCTGATGGCA 669

AUTHORS    Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
           ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
           Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
           Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
           Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
           Sorrells,M., Warburton,M. and Wenzel,G.
TITLE      International Triticeae EST Cooperative (ITEC): Production of
           Expressed Sequence Tags for Species of the Triticeae
JOURNAL    Unpublished (2000)
COMMENT    Contact: Jia J
           Key Lab. of Crop Germplasm & Biotechnology
           Inst. of Crop Germplasm Resources
           Beijing 100081 PR CHINA
           Tel: 86 1 62186623
           Fax: 86 1 62186629
           Email: jizeng@mimi.cnc.ac.cn
           International Triticeae EST Cooperative (ITEC)
           http://wheat.pw.usda.gov/genome.
FEATURES   Location/Qualifiers
     source    1..300
                 /organism="Triticum aestivum"
                 /cultivar="Powdery Mildew Resistant line"
                 /db_xref="taxon:4565"
                 /clone="CNW02EL021"
                 /clone_lib="ITEC CNW Wheat Powdery Mildew Resistant
                 Library"
                 /tissue_type="leaf"
                 /dev_stage="seedling, challenged with powdery mildew
                 strain"
                 /note="Vector: Lambda TripLEX; Site_1: SfiI; Primer used:
                 5'-TCCGAGATCTGGACGAGC-3'  500 bp average insert size."

BASE COUNT       80 a      75 c      67 g      72 t
ORIGIN

Query Match         9.5%;  Score 44.4; DB 167; Length 300;
Best Local Similarity  66.5%;  Pred. No. 0.0038;
Matches   105;  Conservative   0;  Mismatches   50;  Indels    3;  Gaps    3;

Qy    56 ggtccagccgaatccctttatgagcgcttaagtcactgccgccgcttacacgcggactgag 115
         ||.|||||| ||||.|||| ||||||||||||||||||||||||||||||| |||||
Db    30 GGGGCCGTCCCATCGCCTNAGTNGTCGATACATTACTGCCGCGTTTACAACGTCTGCTGG 89

Qy   116 aaacctgctaccggctaatctcttcacacatccccatctgcactgcgcggtatctcaatg 175
         ||||||||| |.|||||||||||| ||| |||||||||||||||||||| ||.||||
Db    90 AAACCTGCGT-TCCACTNATCGCTTGANACATCCCCT-TCGCACTGCGAATAGCGAAAGG 147

Qy   176 accacc-atcgccttccaaccatgccaactaatggcaa 212
         ||| |||||||||||||| ||||| ||||||
Db   148 CCGACCGATCGCCTTCCAACAGGCCANCTGATGGCGA 185

RESULT 10
AQ763115
LOCUS       AQ763115        205 bp    DNA                 GSS        28-JUL-1999
DEFINITION  HS_3161_A2_D02_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate-3161 Col-4 Row=G, DNA sequence.
ACCESSION   AQ763115
VERSION     AQ763115.1  GI:5641231
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 205)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L

---

High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web server: http://www.htsc.washington.edu
Plate: 3161  row: G  column: 4
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 205..
FEATURES   Location/Qualifiers
     source    1..205
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
                 /clone="Plate=3161 Col=4 Row=G"
                 /clone_lib="CIT Approved Human Genomic Sperm Library D"
                 /sex="male"
                 /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                 E-Coli DH10B"
BASE COUNT       41 a      55 c      44 g      65 t
ORIGIN

Query Match         9.4%;    Score 43.6;  DB 233;  Length 205;
Best Local Similarity   80.8%;    Pred. No. 0.0062;
Matches    63;  Conservative    0;  Mismatches   14;  Indels    1;  Gaps    1;

Qy   183 tcgccttccaaccatgccaactaatggcaatggccctgat-cggatttcttcttaccattg 241
         || |||||||| ||| ||||| ||||||||||||||||| | ||| |||||||| |||||||
Db    21 TCCCCTTCCAACAATGTCCATTGATGGCAATGGCCGGATCCGGTTTTCCTCCTACCATTG 80

Qy   242 tcggatttacc\gctatg 259
         |||||||||.|||||||
Db    81 TGTGATATACCCGTTATG 98

RESULT 11
AV721604
LOCUS       AV721604        736 bp    mRNA                EST        16-OCT-2000
DEFINITION  AV721604 HTB Homo sapiens cDNA clone HTBBBEll 5', mRNA sequence.
ACCESSION   AV721604
VERSION     AV721604.1  GI:1082329 6
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 736)
AUTHORS     Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
            ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
TITLE       Homo sapiens cDNA HTB clones
JOURNAL     Unpublished (2000)
COMMENT     Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Contact: Zeguang Han
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: ··h@chgc.sh.cn
            This clo··  · available at CHGC in Shanghai.
FEATURES   ·ion/Qualifiers
     source    ·/36
                 ·ganism="Homo sapiens"
                 ·_xref="taxon:9606"
                 ·lone="HTBBBEll"
                 ·one_lib="HTB"
                 ·tissue_type="Hypothalamus"
                 ··stage="Adult"
                 ··host="SOLR"

## Left column

```
                              /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                              XhoI"
BASE COUNT    207 a    165 c    196 g    167 t    1 others
ORIGIN

Query Match            9.3%;    Score 43.2;  DB 32;  Length 736;
Best Local Similarity  62.9%;   Pred. No. 0.011;
Matches    83;  Conservative    0;  Mismatches    48;  Indels    1;  Gaps    1;

Qy  109  gactgagaaacctgcgtaccgctaatcttcacacatccctatgcactggcggtat  168
         ||||||||||| ||| || |||||| ||||||| ||||||| || ||||| ||||||
Db  541  GACTGGGAAACCTGGCGTACCACTTATCGCCTGAGACATCCCCTTTCGCAGTTGCGTATA 600

Qy  169  ctcaatgaccacc-atcgcttccaaccatgccaactaatggccctgatcggat  227
         ||||| |||| ||| |||||||||||  ||| ||||| ||||  ||| || ||
Db  601  CGACAGGCCGGCGCCATCGCCTTCCACAGTGGCCACCGACCGAATGAAATGAGAGCGTAT  660

Qy  228  ttcctctaccat  239
         ||| |||||||||
Db  661  ATTTGTTCGCGT  672

RESULT 12
LOCUS       AV726956        946 bp       mRNA               EST       17-OCT-2000
DEFINITION  AV726956 HTC Homo sapiens cDNA clone HTCAYD01 5', mRNA sequence.
ACCESSION   AV726956
VERSION     AV726956.1  GI:10836377
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 946)
AUTHORS     Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao.X., Cheng,Z., Xu,Z., Zeng,L., Xu
            ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
TITLE       Homo sapiens cDNA HTC clones
JOURNAL     Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            20120J, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1..946
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HTCAYD01"
                     /clone_lib="HTC"
                     /tissue_type="Hypothalamus"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                              /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                              XhoI"
BASE COUNT    265 a    180 c    220 g    270 t    11 others
ORIGIN

Query Match            9.2%;    Score 43;  DB 32;  Length 946;
Best Local Similarity  62.8%;   Pred. No. 0.014;
Matches    81;  Conservative    0;  Mismatches    47;  Indels    1;  Gaps    1;

Qy  107  cggactgagaaacctgcgtaccgctaatcttcacacatccctatgcgcactggcgggt  166
         |||||||| || |||||||||| ||| ||||| ||||| ||| || ||||||| |||||
Db  609  CGGACTGGAAACCCTGCGTACCCTACCTAATCTGCTTGAGACATCCNCTTTGCAAGTGGCTAT  668

Qy  167  atctcaatgaccaccatcgcttccaaccatgccaactaatggcaatggccctgatcgga  226
```

## Right column

```
Db  669  ACCATAGGCGCGGACGATCGCTTT-CAACATTGNCAGCTGATGCAATTGAAATGTACCGTA  727
         || ||||| |||| || ||||| || ||| | ||| |||   |||||||||| ||||||||
Qy  227  tttcctctta  235
         || ||| | ||
Db  728  TATTTGTAA  736

RESULT 13
AQ823152
LOCUS       AQ823152         168 bp        DNA               GSS       26-AUG-1999
DEFINITION  HS_3186_B1_C05_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3186 Col=9 Row=F, DNA sequence.
ACCESSION   AQ823152
VERSION     AQ823152.1  GI:5785545
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 168)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3186  row: F  column: 9
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 168.
FEATURES             Location/Qualifiers
     source          1..168
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3186 Col=9 Row=F"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
            E-Coli DH10B"
BASE COUNT    40 a    55 c    22 g    49 t    2 others
ORIGIN

Query Match            9.2%;    Score 42.8;  DB 234;  Length 168;
Best Local Similarity  61.9%;   Pred. No. 0.011;
Matches    99;  Conservative    0;  Mismatches    59;  Indels    2;  Gaps    2;

Qy  172  aatgaccaccatcgcttccaaccatgccaactaatggcaatggccctgatcggatttct  231
         || |||||||||||| |||| |||||||| |||||||||||||| ||||| ||||||||||
Db    7  AAGCCCCACCATCGCCTTCCAACATTGCC-ACTAATGGGAATGCCCGATCGGATTTCTC  65

Qy  232  cttacattgtcggattacccgctatggcgcttcatcaaca-gaggatgccctagtatc  290
         || | ||||| | ||| |||||||| ||||||| || |||| | |||| ||| ||||||||
Db   66  CTTCCATCTTGCTCGNGTTTTTTACACGCTGATGTGCCTCTANACATATGTCTATGCCCTATATAC  125

Qy  291  acctaccccacacccgtgaccaaccttggcctttaaata  330
         ||||| |||||| ||| || || ||  ||||||||| |||
Db  126  CACCCAACCGTACCCTTATCCACCATTGTCCATGGGCTCAAATA  165

RESULT 14
AV705287
```

```
LOCUS       AV705287                659 bp    mRNA            EST       09-OCT-2000
DEFINITION  AV705287 ADB Homo sapiens cDNA clone ADBBBE08 5', mRNA sequence.
ACCESSION   AV705287
VERSION     AV705287.1  GI:10722593
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 659)
AUTHORS     Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao
            ,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z.,
            ,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
TITLE       Homo sapiens cDNA ADB clones
JOURNAL     Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1..659
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="ADBBBE08"
                     /clone_lib="ADB"
                     /tissue_type="Adrenal gland"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pbluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      240 a      149 c      128 g      140 t        2 others
ORIGIN

Query Match             9.2%;  Score 42.6;  DB 32;  Length 659;
  Best Local Similarity   74.0%;  Pred. No. 0.017;
  Matches   54;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy         80 cgcttaagtcactgcgccgcttacacggactgagaaacctgctaccgctaatctct 139
              || |  |||||| |||| |||||||||||||   ||  |||||||||||  ||||  |
Db        576 CGATACAATCACTGGCGCGTTTACACGCGGACTGGGAAACCCTGCGTACCAACTTAATGCC 635

Qy        140 tcacacatccccct 152
              | |||||| ||||
Db        636 TGAGACATTCCCT 648


RESULT 15
B45121/c
LOCUS       B45121                  312 bp    DNA             GSS       21-OCT-1997
DEFINITION  HS-1060-B1-G01-MR.abi CIT Human Genomic Sperm Library C Homo
            saplens genomic clone Plate=CT 782 Col=1 Row=N, DNA sequence.
ACCESSION   B45121
VERSION     B45121.1  GI:2549955
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 312)
AUTHORS     Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Tralcoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
TITLE       Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 sequence
            Tagged Connectors
JOURNAL     Unpublished (1997)
COMMENT     Contact: Mahairas GG, Zackrone KD, Hood L
            University of Washington
            Seattle, WA 98195, USA
            Tel: (206) 616-8744
            Fax: (206) 685-7301
            Email: kzackron@u.washington.edu
            Sequence Tagged Connector
            Plate: CT 782  row: N  column: 1
            Class: BAC ends
            High quality sequence stop: 312.
FEATURES             Location/Qualifiers
     source          1..312
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="plate=CT 782 Col=1 Row=N"
                     /clone_lib="CIT Human Genomic Sperm Library C"
                     /sex="M"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT       80 a       66 c       95 g       70 t        1 others
ORIGIN

Query Match             9.1%;  Score 42.4;  DB 256;  Length 312;
  Best Local Similarity   51.9%;  Pred. No. 0.016;
  Matches   94;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        161 ggcggtatccaatgaccacatgcctccaacatgccaactaatggcaatggcccctg 220
              ||  | |  |||||  ||||||||  ||||| |  |  |||||  ||||
Db        226 GGGTTTACCCCGGTGACCGGATTGGCCTTACCCCAAGCCAGTTCCTCGGAAGGATCCTTC 167

Qy        221 atcggattctcttaccattgtcggattaccgctatggcgcttcatcaacagaggatg 280
              | |  |||| ||| || ||   |  ||||||  |  ||  |  ||||||
Db        166 CTAGTTTTCCTCAGAGGATATTCCCTTTTCCACATTAGACCTCATTGTGCACAGAGGATT 107

Qy        281 ccctagtatcaccctacccacacccgtgaccaaccttggcctttaaataaacctatct 340
              ||||  |||| | ||  |  |||  |  |||  |||| |  |||||  |||  |
Db        106 CCCTTCATAGATCCTACAAATACAGTGTNTCCAAGCTGCTCCATTAAAAGAAAGCTTGCA 47

Qy        341 t 341
              |
Db         46 T 46


Search completed: November  7, 2001, 03:48:08
Job time: 1473 sec
```